

**ESSAYS ON CUSTOMER RELATIONSHIP MANAGEMENT IN  
SOCIAL NETWORKS**

A Dissertation

Presented to the Faculty of the Graduate School  
of Cornell University

In Partial Fulfillment of the Requirements for the Degree of  
Doctor of Philosophy

by

Ping Zhao

January 2014

© 2014 Ping Zhao

# ESSAYS ON CUSTOMER RELATIONSHIP MANAGEMENT IN SOCIAL NETWORKS

Ping Zhao, Ph. D.

Cornell University 2014

The past decade has witnessed an explosive growth of social media. It has never been so easy for customers to connect and interact with each other within various forms of social networks. This trend provides companies with unprecedented opportunities to enhance firm-customer relationships by leveraging the power of social influence among customers. However, it also poses potential risks when negative word-of-mouth gets viral on customer networks.

The traditional customer relationship management (CRM) research treats customers independently, which might no longer apply to today's networked customers. Therefore, this dissertation investigates approaches to embed social network analysis components into customer relationship management techniques, thus finding a way to harness the power of social influence to improve the efficiency of companies' CRM efforts.

Essay 1 of this dissertation studies the "group-to-one" social influence of strong and weak ties within a social network, with a framework of a social interaction model, a social influence model, and a tie strength measure. It is found that the social influence mechanism through strong and weak ties is complex. Sharing and reciprocity play an important role in the way social influence affects customers' purchasing behaviors. It is also found that, as a whole, weak ties are more influential

than strong ties.

Essay 2 attempts to model customers' defection decisions within a social network. By jointly estimating a dyadic level tie strength model and an individual level defection decision model, it is found that customers who actively interact with others tend to have strong ties with them. Also, customers with strong ties tend to have stronger influence on other customers' defections than customers with weak ties. In this essay, a new approach to measure customers' social network value is also proposed.

Essay 3 discusses the promising research opportunities in integrating social network analysis (SNA) components into customer relationship management (CRM). It briefly reviews the four critical aspects of CRM: acquisition, retention, growth, and firm-customer relationship dynamics. Within each aspect the discussion focuses on the possible impact of social network components on CRM models, and how to combine CRM and SNA in modeling efforts.

## BIOGRAPHICAL SKETCH

Ping Zhao was born in Jilin and grew up in Shanghai, China. In 1994, Ping got his bachelor's degree in automotive engineering from Tsinghua University in Beijing, China. Since then he had been working for nearly five years at Baosteel Group, as a mechanical engineer and technical administrator. He then moved on to become a data analyst and consultant on management information system. In 2003, Ping went to America for advanced business education. In 2005 he got his M.B.A. degree from the Goizueta Business School of Emory University in Atlanta, Georgia. After graduation he worked as a research analyst at Zyman Institute of Brand Science (ZIBS) for a year. During his days at ZIBS Ping had opportunities to work with renowned marketing faculties on various research projects. With developed interest in marketing science, Ping joined the Ph.D. program at the Johnson School of Cornell University in 2006. In 2013, Ping started working at Wilfrid Laurier University in Waterloo, Ontario Canada, as an assistant professor of marketing.

To my family, for their unconditional love and support

## ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my committee chair, Professor Vithala Rao. Without his guidance, encouragement, patience, and persistent help this dissertation would not have been possible.

I would like to thank my committee members, Professor Martin T. Wells and Professor Vishal Narayan, for their support and help.

I also thank Professor Sriram Venkataraman. His encouragement is the main reason I took the arduous but rewarding path to the Ph.D.

## TABLE OF CONTENTS

Biographical Sketch .....	iii
Dedication .....	iv
Acknowledgements .....	v
Table of Contents .....	vi
List of Figures .....	vii
List of Tables .....	viii
CHAPTER 1 UNDERSTANDING CUSTOMER RELATIONSHIP STRENGTH AND PROFITABILITY IN AN ONLINE SOCIAL NETWORK.....	1
1.1 Introduction .....	2
1.2 Literature Review .....	6
1.3 Data.....	9
1.4 Models .....	15
1.5 Estimation.....	29
1.6 Robustness Tests, Cross-Validation, Causality Inference, and Policy Simulations .....	45
1.7 Discussion.....	51
References .....	54
CHAPTER 2 MODELING CUSTOMERS' DEFECTION IN A SOCIAL NETWORK .....	57
2.1 Introduction .....	58
2.2 Literature Review .....	61
2.3 Data.....	65
2.4 Models .....	69
2.5 Estimation and Results .....	77
2.6 Tests, Cross-Validation, and Policy Simulation.....	81
2.7 Discussion.....	91
Appendix .....	93
References .....	96
CHAPTER 3 CUSTOMER RELATIONSHIP MANAGEMENT AND SOCIAL NETWORK ANALYSIS: POSSIBLE SYNERGY OPPORTUNITIES .....	99
3.1 Introduction .....	100
3.2 CRM and SNA – A Brief Review .....	101
3.3 Acquire New Customers through Social Network and Social Media .....	112
3.4 Leverage the Power of Social Influence to Retain Customers .....	116
3.5 Grow Customers within a Social Network.....	122
3.6 Capture Firm-Customer Relationship Dynamics within a Social Network .....	125
3.7 Conclusion.....	128
References .....	131



## LIST OF FIGURES

Figure 1.1 Monetary Contribution by In-Sample Customers and Out-of-Sample Customers .....	13
Figure 1.2 Sharing and Purchasing Behaviors of a Typical Customer.....	14
Figure 1.3 Research Framework.....	15
Figure 1.4 Cohesion Subgroups .....	17
Figure 1.5 Time-Diminishing Effects.....	20
Figure 1.6 The Identification of Strong/Weak Ties.....	26
Figure 1.7 The Impact of Dyad Duration on Customers' Sharing Decision and Frequency .....	36
Figure 1.8 Comparison of Strong and Weak Ties in Sharing Behaviors .....	38
Figure 1.9 Comparison of the Sizes of Strong Tie Network and Weak Tie Network..	43
Figure 1.10 Policy Simulation Results .....	51
Figure 2.1 The Research Framework .....	69
Figure 2.2 The MCMC Algorithm .....	70
Figure 2.3 Incremental Impacts of Customers' Defection Decisions.....	85
Figure 2.4 Measuring Customers' Social Network Values .....	86
Figure 2.5 Model Can Predict Defection Time with Higher Accuracy.....	89
Figure 2.6 Retention of Influential Customers Can Largely Increase Revenues .....	90

## LIST OF TABLES

Table 1.1 Descriptive Statistics of Sampled Customers.....	12
Table 1.2 Overlap between the Sharing Network and the Friendship Network.....	12
Table 1.3 Models .....	21
Table 1.4 Covariates Used in Models 1 and 2.....	22
Table 1.5 Covariates Used in Models 3 and 4.....	28
Table 1.6 Model Estimation Results .....	30
Table 1.7 Sharing Decision Model (Model 1D) Estimation Results.....	32
Table 1.8 Sharing Frequency Model (Model 2D) Estimation Results .....	34
Table 1.9 Cash Payment Decision Model (Model 3D) Estimation Results .....	39
Table 1.10 Cash Payment Amount Model (Model 4D) Estimation Results .....	41
Table 1.11 Results of Cross-Validation.....	47
Table 1.12 Endogeneity Tests on Subgroup Payment Covariates.....	50
Table 2.1 Descriptive Statistics of Sampled Regular Customers .....	68
Table 2.2 Overlap between the Interacting Network and the Friendship Network.....	68
Table 2.3 Scenarios of Incremental Effects.....	74
Table 2.4 Estimation Results of Tie Strength Model and Defection Decision Model ..	79
Table 2.5 Estimate Results of Benchmark Models .....	83
Table 2.6 Estimation Results of Robustness Test .....	84
Table 2.7 Results of LOO Validation.....	87
Table 3.1 Research on CRM .....	103
Table 3.2 Social Network Analysis (SNA) Research in Marketing .....	107

# CHAPTER 1

## UNDERSTANDING CUSTOMER RELATIONSHIP STRENGTH AND PROFITABILITY IN AN ONLINE SOCIAL NETWORK

### *Abstract*

We study the “group-to-one” social influence of strong and weak ties within a social network in the context of virtual product purchasing. We propose a comprehensive framework combining social interaction models, social influence models, and tie strength measures. We use customer data provided by an online gaming company to test our framework, which is built with four generalized linear mixed models. We find that the social influence mechanism through strong and weak ties is complex. Sharing and reciprocity play an important role in the way social influence affects customers’ purchasing behaviors. We also find that, as a whole, weak ties are more influential than strong ties. We run policy simulations to demonstrate the utility of our framework and discuss how companies can use it to increase their revenues. Our research findings are also applicable to other industries, particularly the eBook and online music industry.

**Keywords** Social Interactions, Social Influence, Tie Strength, Virtual Products, CRM

## 1.1 Introduction

The past decade has witnessed explosive growth within the online video gaming<sup>1</sup> industry in the United States and other countries. In 1999 online video games produced no revenue. Ten years later, *World of Warcraft*, the largest massive multiplayer online game (MMOG), generated nearly \$1.2 billion in revenue for its publisher (Schmidt 2012). Zynga, a social online game company founded in 2007, earned \$850 million in revenue in 2010, increasing more than 200% over the year before (Austin 2011). This young company had already attracted 59 million average daily active users by the second quarter of 2011 (Ovide 2011). According to Schmidt (2012), the revenue from online game subscriptions increased at an annual rate of 40% over the period from 2008 to 2012 and is estimated to reach \$5.9 billion in 2012. Schmidt also projects that the providers of online games will become even more profitable in the future. The same trend is also seen in emerging economies such as China. The PRC's Ministry of Culture (2011) reports that in 2010 China's total online game revenue reached \$5.1 billion, increasing 26.2% from 2009. The number of game players exceeded 120 million in 2010, growing more than 37.0% from 2009.

An important feature of modern games is the social network function that enables players all over the world to socialize and interact through the Internet. Gaming companies painstakingly record customers' social connections and behaviors in databases. These online gaming data provide researchers a precious opportunity to observe customers' behaviors within a social network. Moreover, one "run" of a game on one server usually lasts merely a few months.<sup>2</sup> Thus, a customer's complete lifecycle in a game can be observed, making this an ideal platform for

---

<sup>1</sup> We define "online video game" as a video game played on personal computers without using a console. Multiple players can play together through the Internet.

<sup>2</sup> To attract new business, companies usually shut down the old server after a certain period of time and open a new server so new players can start afresh together.

research on customer relationship management (CRM) within a social network.

Another important feature of modern games is the “free-to-play” business model. Due to intense competition among companies and low switching costs for customers, players can play most online video games free of charge. With this business model, gaming companies have two ways to generate revenues: running advertisements and selling virtual goods to customers for real money. Some companies directly display advertisements in their games. Some promote other companies’ applications in the games and charge for the number of installs and clicks (Chang and Mendelson 2010). Recently, selling virtual goods has become the most important revenue source for gaming companies. According to Greengard (2011), in 2010 virtual goods sales generated 60% of revenues, and “the demand for virtual goods hit \$7.3 billion in 2010, up from \$2.1 billion in 2007.... [T]he figure will reach \$14 billion in 2014” (p. 19).

Even with the huge and fast-growing demand for online games, gaming companies find it hard to earn customers’ money. Convincing customers to pay real money for virtual products that can only be used within the game is a challenge. The gaming industry has a well-known “95-5” rule: only about 5% of the customers actually purchase virtual goods with real money; the remaining 95% never pay a nickel. Therefore, building a large customer base and motivating customers to pay more is critical for companies’ profitability. Consequently, some leading companies have started collecting and analyzing customer data to find ways to increase profits. For example, behind Zynga’s impressive success is its relentless effort to analyze and understand its customers’ virtual product preferences. Zynga’s vice president in charge of its data analysis team said it best: “We’re an analytics company masquerading as a games company” (Wingfield 2011, para. 5).

Knowledge of how customers influence each other’s purchasing behavior within the

game is still very limited. To generate more profits, gaming companies need to understand the factors that affect their customers' spending habits, especially high-value customers. Given the social network nature of these games, social connections, interactions, and influence might all play important roles in the value-contribution process.

In this study, we attempt to answer the following two questions: (1) How can we identify strong and weak ties within a social network, based on customer social connection and activity information? (2) Do social influence patterns differ among customers with strong and weak ties? To answer these questions, we develop a comprehensive framework that combines social interaction models, social influence models, and tie strength measures.

Our framework is built with four generalized linear mixed models (GLMMs), which we test using customer data across a four-month period provided by a gaming company in China. We first use social interaction models to analyze customers' sharing behaviors, from which we uncover the factors that influence their relationship strength. Based on these findings, we calculate directional, continuous, and time-varying tie strength measures for all customer dyads in each week. With this tie strength information, we are able to remap the social network as a valued, directional graph evolving over time. For every week, we categorize each customer's active social ties into four cohesion subgroups based on the direction and strength of the relationship. Finally we use social influence models to investigate the aggregate social influence of cohesion subgroups on an individual customer's purchasing decisions.

To the best of our knowledge, our research is the first to analyze the "group-to-one" social influence through strong and weak ties. This study is also the first to model social influence in the context of virtual product purchasing. We find that the social influence mechanism of strong ties and weak ties is complex. Sharing and reciprocity play an important

role in the way social influence affects customers' purchasing behaviors. We also find that weak ties, as a whole, are more influential than strong ties.

This research also illustrates that “behavioral” network characteristics outperform “nominal” network characteristics in explaining the variation in customer's social interactions and monetary contribution.<sup>3</sup> We find that actively interacting customers have stronger influence on each other than customers who merely register as friends or who are members of the same online groups.

To demonstrate the benefits of our framework, we use two policy simulations to showcase how our models can help increase a gaming company's revenue. We find that by increasing actively interacting dyads by 10% or increasing customers' online time by 10%, the gaming company can boost its revenue by more than 40%.

The benefits of our framework go beyond the gaming industry. Our findings shed light on customers' social connections, interactions, and influences in the physical world. Prior research has shown that customers have similar motivations and behaviors in the virtual world of games and in the physical world of their daily lives (Bartle 2003; Malone 1981; Williams, Yee, and Caplan 2008). Researchers also have found that people tend to treat virtual products and physical products in similar ways and that the basic laws of supply-and-demand still apply. This explains why people are willing to pay real money for virtual products (Greengard 2011).<sup>4</sup> Hence, our findings can help companies improve their CRM efficiency in the physical world by leveraging the power of relationships in a social network.

Our findings are particularly relevant to the marketing of non-gaming digital products

---

<sup>3</sup> “Behavioral” network characters are calculated from customers' social interaction records, while “nominal” network characteristics are calculated from customers' self-inputted registration records. Detailed explanation of these two concepts is given in the “Models” section.

<sup>4</sup> In fact, recently some companies have begun implementing multiplayer online games to train employees (Reeves and Read 2009) or motivate people to solve the real-world problems (McGonigal 2011).

such as eBooks and MP3 music tracks. Fairfield (2005) and Lehonvirta (2009) point out that virtual gaming products and non-gaming digital products are different in terms of the change of ownership that takes place during transactions. For instance, when a gamer gives or sells a virtual product to another gamer, she loses ownership of this product. However, when a customer gives or sells an MP3 track to another customer, she still owns the track after the transaction. Aside from this ownership issue, these two categories are similar in terms of product distribution and consumption. Especially for products for which sharing is allowed, our research can provide valuable insights into consumer purchases of these non-gaming digital products within a social network.<sup>5</sup>

The remainder of this paper is organized as follows: in section 2 we review the relevant literature; then in section 3 we describe the data used in this research. Section 4 discusses the model specification. The estimation results are presented in section 5. In section 6 we run the robustness tests, cross-validations, causality inference, and policy simulations. In section 7 we conclude this paper with a discussion of the contributions and limitations of our research, and future research opportunities.

## **1.2 Literature Review**

This literature review focuses on three streams of literature around which we build our comprehensive framework: social interaction, social influence, and tie strength.

Research on social interactions has a long history in marketing, sociology, and computer science. Most of this literature concentrates on the identification of social connections. For example, Iacobucci (1990) discusses various methods to identify social subgroups from sociometric data. Hoff, Raftery, and Handcock (2002) use a latent variable model to predict the

---

<sup>5</sup> Barnes & Noble's Nook and Amazon's Kindle now allow their customers to share books with friends (but Apple still forbids the sharing of eBooks from the Apple Store).



existence of relational ties. Ansari, Koenigsberg, and Stahl (2011) develop a framework to simultaneously model multiple relationships. Shriver, Nair, and Hofstetter (2011) discuss the self-reinforcing effects between the formation of social ties and the posting of self-generated content. Hartmann et al. (2008) discuss the modeling of active and passive social interactions.

Literature on social influence is also rich. For instance, Nair, Manchanda, and Bhatia (2010) use linear models to investigate the asymmetric social influence of opinion leaders on physicians' prescription behaviors. The authors detail their methods of inferring causality (peer effects) from the correlation of these behaviors. Christakis and Fowler (2007) use longitudinal models to analyze the spread of obesity through a social network. Trusov, Bodapati, and Bucklin (2010) use a Poisson regression model and a Bayesian shrinkage algorithm to analyze customers' influence on each other's log-on behaviors and to identify influential customers. Ma, Krishnan, and Montgomery (2011) use a hierarchical Bayesian model to investigate the impact of homophily<sup>6</sup> and social influence on customers' purchasing timing and product choice decisions. Narayan, Rao, and Saunders (2011) use a two-stage conjoint-based approach to model peer effects on customers' preferences for product attributes in a social network. Leenders (2002) models social influence through network autocorrelation. Hartmann et al. (2008) discuss spillover and multiplier effects in social influence.

A large portion of the tie strength literature focuses on the structural role of strong ties and weak ties in a social network. In his seminal paper, Granovetter (1973) elaborates the importance of weak ties. He finds that weak ties can serve as the "bridges" connecting isolated communities, which cluster together through strong ties. Recently, with increasingly available online social network data, researchers have found evidence to support Granovetter's weak tie

---

<sup>6</sup> In social science, homophily (McPherson, Smith-Lovin, and Cook 2001) usually refers to the fact that people with similar characteristics tend to form strong ties.

theory. For instance, Ferrar et al. (2012) use Facebook data to quantitatively assess the strength of weak ties. They find that weak ties provide better access to information and opportunities. Bakshy et al. (2012) run a large-scale field experiment on Facebook and find that weak ties are important for the spread of online information. Some researchers, on the other hand, have found that strong ties can be important under certain circumstances. For example, Farrow and Yuan (2011) find that tie strength has an influence on people's attitudes and behaviors in the context of volunteerism and charity.

Most researchers identify strong and weak ties using simple metrics (Marsden and Campbell 1984). These metrics include time spent together (Granovetter 1973), communication frequency (Bakshy et al. 2012; Farrow and Yuan 2011), or "emotional closeness" determined through surveys (Farrow and Yuan 2011). Some researchers have adopted a modeling approach to measure tie strength as a function of parameters. For example, Iacobucci and Hopkins (1992) apply a log-linear model to a  $Y$ -array to estimate discrete level tie strength. With this technique, the authors predict the probability that actor  $i$  is related to  $j$  with strength  $k$  and  $j$  related to  $i$  with strength  $l$ . Xiang, Neville, and Rogati (2010) propose a hierarchical model to obtain a latent, continuous tie strength measure. In their model, latent tie strength is constructed as a function of homophily parameters (e.g., living at the same location, going to the same school, the number of common friends) and is estimated from people's interactions (e.g., communication, tagging).

We found several relevant gaps in the literature. First, "group-to-one" social influence is an area that remains largely unexplored. Most of the social influence literature concentrates on "one-to-one" social influence (e.g., physician-physician, blogger-reader). However, such one-to-one social influence might not be applicable in the context of virtual product purchasing, where an individual customer is usually immersed in the influence of a group of customers with whom

she interacts.

Second, most of the literature does not incorporate the direction and strength of social relationships when measuring social influence. It remains unclear whether the social influence mechanism operates differently through strong ties and weak ties.

Finally, the measurement of relationship strength has room for improvement. Using gaming data, we are able to develop a directional, continuous, and time-varying tie strength measure. In other words, this measure should contain as much information about the nature of the relationship as possible. It should also be the function of various relevant factors and should therefore enable us to understand the how these factors influence a customers' relationships.

To fill these gaps in the literature, we develop a comprehensive framework combining social interaction models, social influence models, and new measures of tie strength. We then use this framework to investigate group-to-one influence through strong ties and weak ties in the context of virtual product purchasing behavior.

### **1.3 Data**

#### *Data Source*

The data used in this research were provided by an online video game company in China. The company records every customer's registration information and activity information. The registration information includes *virtual* demographic and social connection information. Two types of social connections exist in most online games: friendship and guild membership.<sup>7</sup> Customers become friends through an "invitation-confirmation" process (similar to Facebook). Any customer can form an online group called a "guild" and become its administrator.

To advance in the game, customers must participate in a variety of activities to earn game

---

<sup>7</sup> "Guild" is gaming jargon referring to online groups that customers form voluntarily.

money and obtain virtual goods. Most items can be purchased from virtual merchants or from other customers with game money.<sup>8</sup> Special items (especially powerful items) can only be purchased with real cash.<sup>9</sup> Customers have limited storage space in the game. If customers have more items than they can carry, they have three options to dispose of them: (1) sell the items to virtual merchants and get 25% of the items' labeled value; (2) sell them to other customers; or (3) simply give away the items free of charge. Interestingly, according to the data, most of the exchanges (more than 95%) are free. Another interesting finding is that most customers naturally strike a balance between their giving and receiving activities.

### *Observation Period*

Our data includes customer activities between February 22 and June 21, 2011 (exactly four months).<sup>10</sup> The game, however, was launched at the end of December 2010. In other words, for customers who registered before February 22 and were still active during our observation period, we do not have a record of their activities prior to that day. This left truncation issue applies to about 39% of the customers and 22% of the dyads<sup>11</sup> in our final sample. The server was closed after June 21. Therefore, we are able to observe the true ending of all dyadic relations, and there is no right truncation issue.

### *Multiple IDs*

No rule prevents the use of multiple IDs in a game. In fact this is a common practice among experienced game players. Customers use multiple IDs for three main reasons: (1) to gain more storage space, (2) to facilitate advancement in the game, and (3) to “gold dig.” In games

---

<sup>8</sup> Game money circulates in the game and cannot be exchanged for real cash.

<sup>9</sup> Games are usually easy and absolutely free at the beginning. As the games progress, they become more and more challenging. After a certain stage, if players want to advance, they need particular powerful items that can only be purchased with cash. Without these items, customers can still advance, but it takes much more time and effort.

<sup>10</sup> Game data can be overwhelming because the system records all activities of all customers (accurate down to a second). The system of this game automatically backed up the data on a four-month rolling basis.

<sup>11</sup> In these dyads, both customers are truncated.

(especially the popular ones) groups of professional players (a.k.a. “gold diggers”) sometimes try to make real money from the game. They usually work in teams to obtain hard-to-get items and sell them to other players, under the table, for cash. These gold diggers tend to use multiple IDs to fetch valuable items and cover their traces. (Gold digging is forbidden in games.)

Multiple IDs could contaminate the data and bias estimates. For instance, two customers who are identified as strongly tied might in fact be two IDs controlled by the same customer. To eliminate this problem, we detect multiple IDs by checking registration information and activity records.<sup>12</sup> We then merge the multiple IDs held by the same customer and give this customer a new unique ID.

### *Sampling the Social Network*

We construct our sample using two criteria. First, sampled customers must be regular customers (i.e., these customers must have log-in records documenting time spent in the game across a period of at least 30 days). Second, sampled customers must have interacted actively with others in the game (i.e., interacted with at least one other customer for at least four weeks). Our sample contains 181 customers, 2,524 directional sharing dyads, 1,315 non-directional pairs of friends, and 42 guilds. Table 1.1 presents descriptive statistics of this social network. In Table 1.2 we cross-tabulate membership in the friendship network against the sharing dyads network and find only a small overlap between these two networks. Thus, many players who registered as friends actually did not interact at all. Similarly, many customers who did interact were not friends. Thus, we include both friendship and interaction covariates in the model.

---

<sup>12</sup> Most of the multiple ID holders are easy to pinpoint, because they either use the same user ID for multiple user names, or use IDs and names with similar combinations of letters and numbers. We also double-checked the activity records of some “suspects.” We doubt that professional gold diggers would target a small game like this one and then would use sophisticated approaches to cover their traces.

**Table 1.1**  
**Descriptive Statistics of Sampled Customers**

Variable	min	max	mean	median	std dev
Number of Friends	0.00	123.00	25.01	19.00	23.60
Number Guilds Joined <sup>13</sup>	0.00	11.00	2.72	2.00	2.16
Number of Sharing Dyads (In-Degree)	0.00	27.00	2.22	1.00	2.79
Number of Sharing Dyads (Out-Degree)	0.00	21.00	2.22	1.00	2.67
Online Time (Hours per Week)	0.00	167.55	31.89	13.09	41.71
Number of Missions Accomplished per Week	0.00	533.00	19.17	1.00	47.35
Tenure (Weeks in Game since Registration)	3.00	24.00	13.62	13.00	6.34
Dyad Duration (Week)	1.00	18.00	3.24	1.00	3.51
Weekly Sharing Frequency (of Sharing Customers)	1	472	4.64	2	11.70
Weekly Cash Payment (of Paying Customers, RMB) <sup>14</sup>	2	13,064	745.50	294	1,567.33

**Table 1.2**  
**Overlap between the Sharing Network and the Friendship Network**

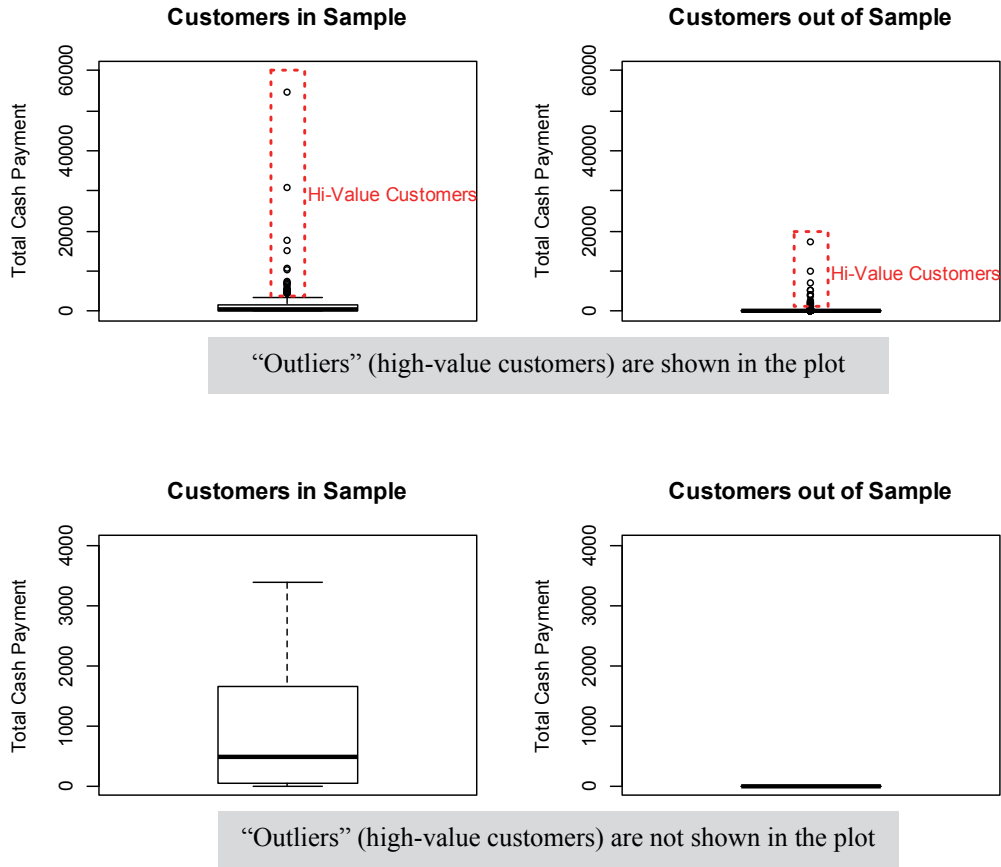
Friendship Relationship	Sharing Relationship		
	No	Yes	Sub Total
No	0 0.00%	669 33.72%	669 33.72%
Yes	722 36.39%	593 29.89%	1,315 66.28%
Sub Total	722 36.39%	1,262 63.61%	1,984 100.00%

To ensure that the sampled customers are also valuable customers, we compare the payments made in real money by the customers in the sample and by those who played the game but were not included in our sample (hereafter, “out-of-sample”) (Figure 1.1).

<sup>13</sup> We deleted all of the single-member guilds.

<sup>14</sup> 1RMB  $\approx$  0.16USD

**Figure 1.1**  
**Monetary Contribution by In-Sample Customers and Out-of-Sample Customers**



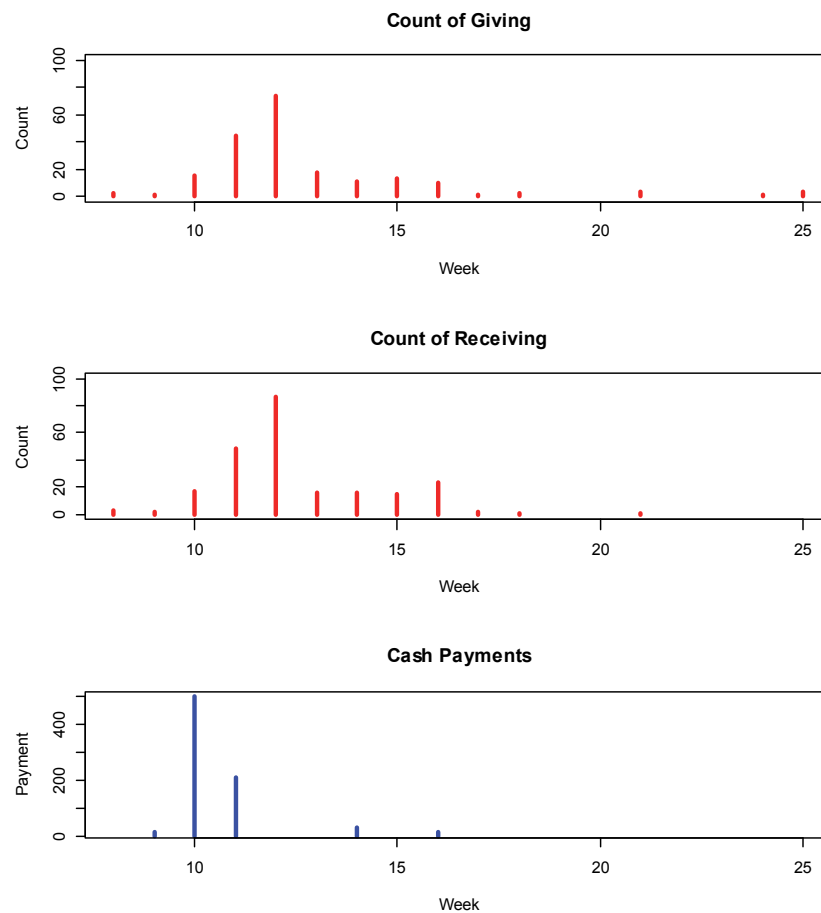
These box plots show that, on average, sampled customers made much higher contributions than the customers who were not sampled. The t-test shows that this difference is statistically significant: the mean payment of the sampled customers is \$289.50 higher than that made by out-of-sample customers ( $p < 0.001$ ).

### *Zero-Inflated Data*

Sharing and payment observations are usually sparse in gaming data. In our sample, sharing did not happen in 48.01% of the dyad-week sharing observations. Payment observations are even less common: customers made no payments in 76.44% of the customer-week observations.

Note that our sample consists of active customers only (playing the game for at least 30 days and interacting with at least one other customer for at least four weeks). If we were to look at the whole customer population, these two figures would be much lower (95-5 rule). Figure 1.2 illustrates the sharing and payment pattern of a typical customer.

**Figure 1.2**  
**Sharing and Purchasing Behaviors of a Typical Customer**



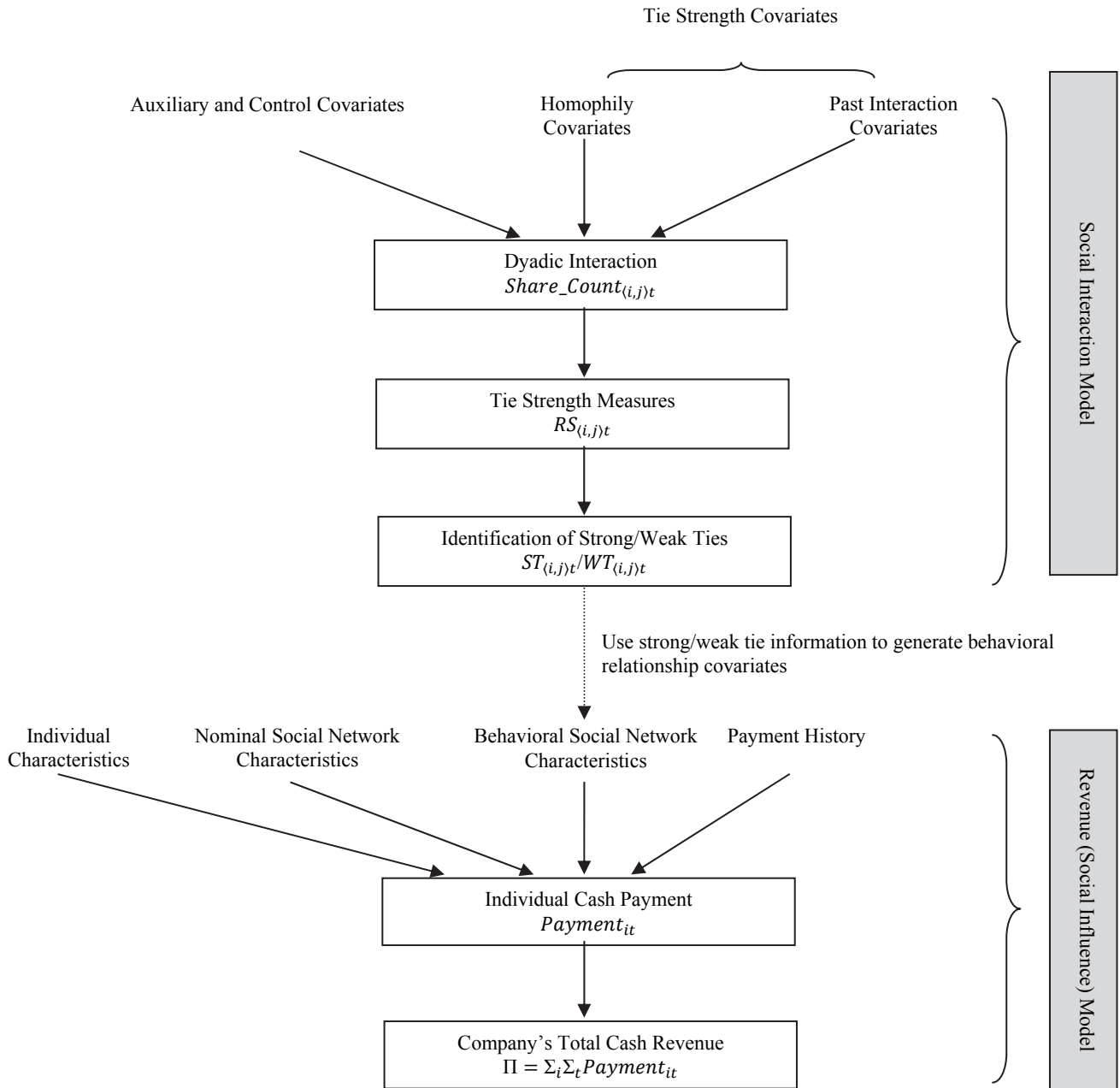


## 1.4 Models

### Framework

Our framework has two interrelated components: a dyadic-level social interaction model and an individual-level revenue (social influence) model (Figure 1.3).

**Figure 1.3**  
**Research Framework**



Customers' cash payments are the most important revenue source for a gaming company. A company's cash revenue from a game is the sum of all its customers' cash payments across their tenures in the game:

$$\Pi = \sum_{i=1}^N \sum_{t=T_i^O}^{T_i^F} \pi_{it}. \quad (1.1)$$

$\Pi$  is the company's total cash revenue from the game, and  $\pi_{it}$  is the cash payment made by customer  $i$  in week  $t$ .  $N$  is the total number of customers.  $T_i^O$  is the starting time of customer  $i$ 's tenure, and  $T_i^F$  is the ending time of customer  $i$ 's tenure.

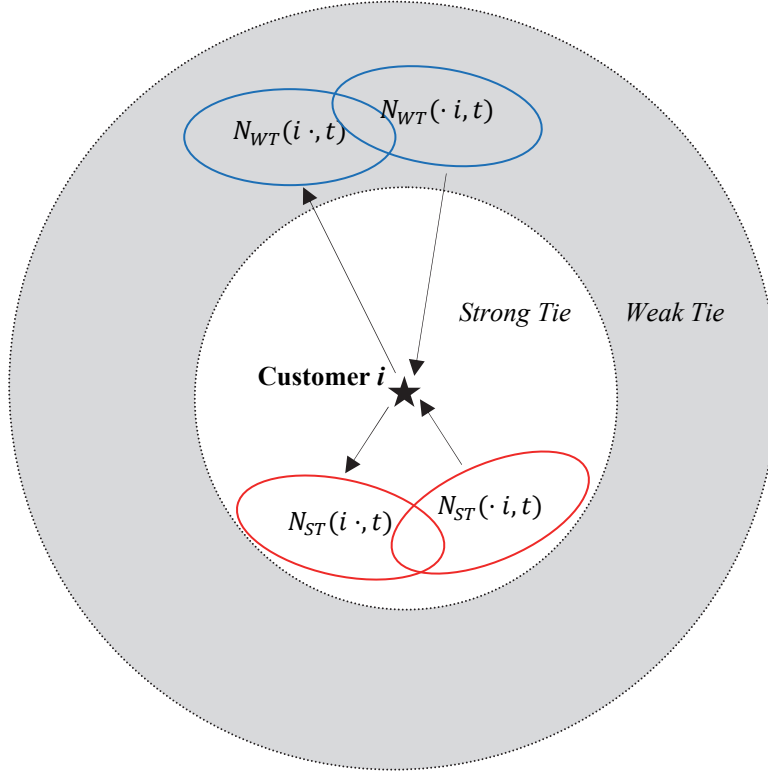
In our *social influence model*, customer  $i$ 's cash payment in week  $t$ ,  $\pi_{it}$  is described as the function of her individual characteristics, her social network characteristics, and her payment history. One novelty in this model is that we try to capture the group-to-one social influence of both strong ties and weak ties on her purchasing behaviors.

In our *social interaction model*, we use customers' sharing behaviors to gauge the strength of their relationships. We make this decision for three reasons: (1) Sharing is the true interaction between two customers and is well recorded; (2) sharing is directional, and the direction of sharing contains information about the relationship, such as favor and prestige (Wasserman and Faust 1994); and (3) sharing imposes opportunity costs on the givers. Giving an item to another customer for free means giving up 25% of the label price, which could be easily earned by selling this item to a virtual merchant. Thus, frequent sharing is an indicator of a close and friendly relationship.

We use *tie strength measures* as a link between our social interaction and social influence models. With estimated coefficients of the social interaction model, we can calculate the tie strength within each dyad in each week. Using these measures, we are able to identify all strong ties and weak ties. Therefore, for each focal customer  $i$ , we cluster her active social ties into four

cohesion subgroups based on the direction and strength of the ties (Figure 1.4).

**Figure 1.4**  
**Cohesion Subgroups**



Direction \ Strength	Strong	Weak
	Strong	Weak
From $i$	$N_{ST}(i, t)$	$N_{WT}(i, t)$
To $i$	$N_{ST}(\cdot, i, t)$	$N_{WT}(\cdot, i, t)$

We then generate various aggregate covariates of these four subgroups, which we then use in the social influence model to capture the group-to-one influence.

### *Modeling Issues*

Three modeling issues must be addressed in our research: zero-inflated data, dyadic interdependence, and temporal correlation.

#### (1) Modeling zero-inflated data

In our sample 48.01% of dyad-week sharing observations are 0, and 76.44% of customer-week payment observations are 0. In such a case, even models with zero-inflated distributions (e.g., zero-inflated Poisson / negative binomial models) might not be sufficient. Therefore, we

adopt a two-stage modeling approach (Ansari, Koenigsberg, and Stahl 2011; Gelman and Hill 2007).<sup>15</sup> First, we model the dyads’ sharing decisions (binary) with a logit model; then, for those dyads that do share items, we analyze their weekly sharing frequency using a Poisson regression model.<sup>16</sup> Similarly, we analyze customers’ payment decisions with a logit model; then, for customers who do make payments, we analyze their payment amounts with a lognormal regression model. In addition to its convenience in modeling zero-inflated data, this approach also gives us more flexibility. We can use different sets of covariates to separately model the decisions and the counts (amounts).

## (2) Modeling dyadic interdependence

Dyadic interdependence is a critical issue in social network modeling. Obviously, the interactions in dyad  $\langle i, j \rangle$  and  $\langle j, k \rangle$  are not independent, because these two dyads share the same customer  $j$ . Therefore their error terms in the models are correlated. To address this issue, we use a generalized linear mixed model (GLMM) with “sender” and “receiver” random effects (intercepts). Conditional on these random effects, the likelihood of dyads can be considered independent (Hoff 2003; McCulloch and Searle 2001).

Formally, in the dyadic interaction case, if we define  $Y_{\langle i, j \rangle}$  as the random variable of interaction from customer  $i$  to customer  $j$  ( $Y_{\langle i, j \rangle}$  could be a binary or a count variable),  $\gamma_{\langle i, j \rangle} = a_i + b_j + \epsilon_{ij}$ ,  $\epsilon_{ij} \sim N(0, \sigma_{\gamma 1}^2)$  as the random effect with  $i$  as “sender” and  $j$  as “receiver”, and  $f_Y$  as the density of  $Y_{\langle i, j \rangle}$ , then—conditional on random effect  $\gamma_{\langle i, j \rangle}$ —the distribution of  $Y_{\langle i, j \rangle}$  could be considered independent (i.e., conditional independence):

$$y_{\langle i, j \rangle} | \gamma_{\langle i, j \rangle} \sim \text{indep. } f_{Y_{\langle i, j \rangle} | \gamma_{\langle i, j \rangle}}(y_{\langle i, j \rangle} | \gamma_{\langle i, j \rangle}). \quad (1.2)$$

<sup>15</sup> Statisticians also call it “two-part” model (Duan et al. 1983).

<sup>16</sup> We also tried using a negative binomial distribution (NBD) model; we found that it returns very similar results, but the algorithm takes a much longer time to converge. Trusov, Bodapati, and Bucklin (2010) have a similar finding.

When modeling individual customer's payments, we still face the interdependence issue. The error terms of customers who are involved in dyadic relations might be correlated as well. In a similar vein, we use GLMM with individual random effects (intercepts) to control for such dependence. Hence, we model individual customer's payments as

$$y_i | \gamma_i \sim \text{indep. } f_{Y_i | \gamma_i}(y_i | \gamma_i), \quad (1.3)$$

where  $Y_i$  is the payment variable of customer  $i$  (payment decision or payment amount), and  $\gamma_i = a_i + \epsilon_i, \epsilon_i \sim N(0, \sigma_{\gamma^2})$  is the random effect term of customer  $i$ .

### (3) Modeling temporal correlation

Individuals and dyads in our data are measured repeatedly during their stay in the game. Therefore, the observations could be correlated across time. To address this issue, we introduce cumulative terms with time-diminishing effects into all four models, following Guadagni and Little (1983) and Trusov, Bodapati, and Bucklin (2010). For dyadic interactions, we define the cumulative interaction term at time  $t$  as

$$X_{\langle i, j \rangle t}^P = \sum_{d=1}^D X_{\langle i, j \rangle t-d} \cdot \theta(d, \rho), \quad (1.4)$$

where  $\theta(d, \rho)$  is the weight;  $d$  is the time lag; and  $\rho$  is the time-diminishing effect parameter.

$X_{\langle i, j \rangle t}$  is dyad  $\langle i, j \rangle$ 's interaction (decision/count) at time  $t$ , and  $D$  is the width of the time window.

Similarly, the cumulative term for the individual customer's payment decision and payment amount are calculated by

$$X_{it}^P = \sum_{d=1}^D X_{i, t-d} \cdot \theta(d, \rho), \quad (1.5)$$

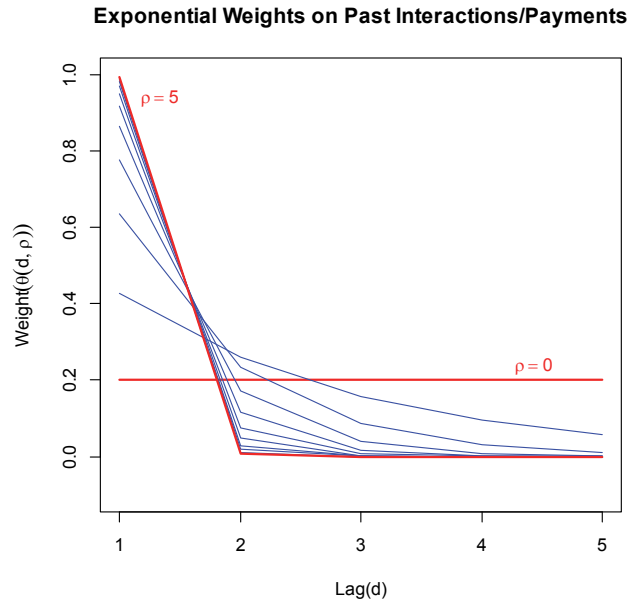
where  $X_{it}$  is customer  $i$ 's payment decision/amount at time  $t$ .

We define the time window as  $D = 5$ , because most of the dyadic relations lasted five weeks or shorter. The weight  $\theta(d, \rho)$  is calculated by

$$\theta(d, \rho) = \frac{\exp(-d \cdot \rho)}{\sum_{l=1}^D \exp(-l \cdot \rho)}. \quad (1.6)$$

Note that  $\sum_{d=1}^D \theta(d, \rho) = 1$ .  $\rho$  is a model parameter to be estimated. Different  $\rho$  values yield different schemes to weigh the past interactions or payments. From Figure 1.5 we see that when  $\rho = 0$ , we get  $\theta(1, \rho) = \dots = \theta(D, \rho) = 1/D$ . In this case, customers have a “long memory” about their past activities, and past activities have equal impact on their current activities. When  $\rho \geq 5.0$ , we have  $\theta(1, \rho) \approx 1, \theta(2, \rho) \approx 0, \dots, \theta(D, \rho) \approx 0$ , which means that customers have a “short memory”: only the most recent activities affect their current activities.

**Figure 1.5**  
**Time-Diminishing Effects**



To simplify the estimation, instead of estimating  $\rho$  jointly with other parameters, we estimate the model over a grid of  $\rho$  from 0 to 5. This is possible because if the  $\rho$  value changes, it only affects the value of this cumulative term, not other covariates. We draw the grid with  $\rho$  from 0 to 5 and calculate the value of that cumulative term at each point of the grid. We choose the optimal  $\rho$  value that yields the best model fit according to log likelihood, Akaike information criterion (AIC), and Bayesian information criterion (BIC) measures.

### Model Specification

We develop four models ( $m = 1, \dots, 4$ ) in our framework. All four models are generalized linear mixed models (GLMMs) with different types of distributions and random effects terms (Table 1.3).

**Table 1.3**  
**Models**

Model	Domain	Outcome	Scale	Distribution
1	Sharing	Sharing decision $y_{\langle i,j \rangle t}^S = \begin{cases} 1, & \text{share} \\ 0, & \text{o. w.} \end{cases}$	Binary	Logistic
2	Sharing	Weekly sharing frequency $y_{\langle i,j \rangle t}^{SF}$	Count	Poisson
3	Payment	Payment decision $y_{it}^P = \begin{cases} 1, & \text{pay} \\ 0, & \text{o. w.} \end{cases}$	Binary	Logistic
4	Payment	Payment amount $y_{it}^{PA}$	Continuous ( $\geq 0$ )	Lognormal

#### Model 1 - Sharing Decision Model

We model dyad  $\langle i, j \rangle$ 's sharing decision ( $i$ 's decision to give items to  $j$ ) with a logit model. We define customer  $i$ 's utility gained from her giving items to customer  $j$  as

$$u_{\langle i,j \rangle t}^S = \underbrace{\mathbf{X}_{\langle i,j \rangle t}^{A[1]'} \cdot \boldsymbol{\beta}^{A[1]} + \mathbf{X}_{\langle i,j \rangle t}^{H[1]'} \cdot \boldsymbol{\beta}^{H[1]} + \mathbf{X}_{\langle i,j \rangle t}^{P[1]'} \cdot \boldsymbol{\beta}^{P[1]}}_{\text{fixed effects}} + \underbrace{a_i^{[1]} + b_j^{[1]}}_{\text{random effects}} + \underbrace{\epsilon_{ijt}^{[1]}}_{\text{error}} \quad (1.7)$$

The sharing decision is

$$y_{\langle i,j \rangle t}^S = \begin{cases} 1, & u_{\langle i,j \rangle t}^S > 0 \\ 0, & \text{otherwise} . \end{cases}$$

The covariates used in this model are listed in Table 1.4.

**Table 1.4**  
**Covariates Used in Models 1 and 2**

Covariates		Type	Model		Description
			1	2	
$X^A$	$Dyad\_Count_{i,t}$	Generosity	x		The number of customers <i>excluding j</i> who receive items from <i>i</i> in week <i>t</i>
	$Share\_Count_{i,t}$	Generosity		x	The number of times <i>i</i> gives items to customers <i>excluding j</i> in week <i>t</i>
	$Dyad\_Count_{j,t}$	Generosity	x		The number of customers <i>excluding i</i> to whom <i>j</i> gives items in week <i>t</i>
	$Share\_Count_{j,t}$	Generosity		x	The number of times <i>j</i> gives items to customers <i>excluding i</i> in week <i>t</i>
	$Dyad\_Count_{\cdot,i,t}$	Popularity	x		The number of customers <i>excluding j</i> from whom <i>i</i> gets items in week <i>t</i>
	$Share\_Count_{\cdot,i,t}$	Popularity		x	The number of times <i>i</i> receives items from customers <i>excluding j</i> in week <i>t</i>
	$Dyad\_Count_{\cdot,j,t}$	Popularity	x		The number of customers <i>excluding i</i> from whom <i>j</i> gets items in week <i>t</i>
	$Share\_Count_{\cdot,j,t}$	Popularity		x	The number of times <i>j</i> receives items from customers <i>excluding i</i> in week <i>t</i>
	$\log(Fight\_Count_{it})$	Devotion	x	x	The log count of customer <i>i</i> 's "monster fights" in week <i>t</i>
	$\log(Fight\_Count_{jt})$	Devotion	x	x	The log count of customer <i>j</i> 's "monster fights" in week <i>t</i>
	$Last\_2W$	Event	x	x	The indicator of week 24 and week 25 of 2011
$X^H$	$Friend\_Count_{it}$	Nominal network	x	x	The number of friends <i>i</i> has in week <i>t</i>
	$Friend\_Count_{jt}$	Nominal network	x	x	The number of friends <i>j</i> has in week <i>t</i>
	$Common\_Friend\_Count_{ijt}$	Nominal network	x	x	The number of <i>i</i> and <i>j</i> 's common friends in week <i>t</i>
	$Guild\_Count_{it}$	Nominal network	x	x	The number of guilds that <i>i</i> has joined in week <i>t</i>
	$Guild\_Count_{jt}$	Nominal network	x	x	The number of guilds that <i>j</i> has joined in week <i>t</i>
	$Common\_Guild\_Count_{ijt}$	Nominal network	x	x	The number of guilds that <i>i</i> and <i>j</i> have joined together in week <i>t</i>
	$Guild\_Leader_{it}$	Nominal network	x	x	The indicator of whether <i>i</i> is the leader of any guilds in week <i>t</i>
	$Guild\_Leader_{jt}$	Nominal network	x	x	The indicator of whether <i>j</i> is the leader of any guilds in week <i>t</i>
	$Dyad\_Duration_{ijt}$	Behavioral network	x	x	The number of weeks since <i>i</i> and <i>j</i> 's first interaction
	$Dyad\_Duration_{ijt}^2$	Behavioral network	x	x	The squared term of dyad duration
	$Tenure_{it}$	Behavioral network		x	Customer <i>i</i> 's tenure in week <i>t</i>
	$Tenure_{jt}$	Behavioral network		x	Customer <i>j</i> 's tenure in week <i>t</i>
	$Tenure\_Diff_{ij}$	Behavioral network	x		The gap between <i>i</i> and <i>j</i> 's tenures ( $Tenure_{it} - Tenure_{jt}$ )
	$Tenure\_Avg_{ijt}$	Behavioral network	x		The average of <i>i</i> and <i>j</i> 's tenures in week <i>t</i> ( $(Tenure_{it} + Tenure_{jt})/2$ )
	$Level_{it}$	Behavioral network		x	Customer <i>i</i> 's level in week <i>t</i>
	$Level_{jt}$	Behavioral network		x	Customer <i>j</i> 's level in week <i>t</i>
	$Online\_Time\_Overlap\_Pcnt_{ijt}$	Behavioral network	x	x	The % of time when <i>i</i> and <i>j</i> stay in the game simultaneously in week <i>t</i>
	$PK\_Count_{ijt}$	Behavioral network	x	x	The number of "personal killings" that <i>i</i> and <i>j</i> finished in week <i>t</i>
	$Mission\_Similarity_{ijt}$	Behavioral network		x	The % of <i>unique</i> missions customer <i>i</i> and <i>j</i> participate together in week <i>t</i>



**Table 1.4 (Continued)**

Covariates		Type	Model		Description
			1	2	
$X^p$	$Cum\_Share\_WK_{ijt}(\rho^S)$	Interaction history	x		The cumulative number of weeks in which $i$ gives items to $j$ during the past five weeks, calculated with time-diminishing parameter $\rho^S$
	$Cum\_Share\_WK_{jit}(\rho_R^S)$	Interaction history	x		The cumulative number of weeks in which $j$ gives items to $i$ during the past five weeks, calculated with time-diminishing parameter $\rho_R^S$
	$Cum\_Share\_Count_{ijt}(\rho^{SF})$	Interaction history		x	The cumulative number of times $i$ gives items to $j$ during the past five weeks, calculated with time-diminishing parameter $\rho^{SF}$
	$Cum\_Share\_Count_{jit}(\rho_R^{SF})$	Interaction history		x	The cumulative number of times $i$ receives items from $j$ during the past five weeks, calculated with time-diminishing parameter $\rho_R^{SF}$

$X_{ijt}^{A[1]}$  is the vector of auxiliary covariates, which includes two generosity covariates, two popularity covariates, two devotion covariates, and one event covariate. We use these auxiliary covariates to control for the factors that have an impact on customers' interactions but have nothing to do with their relationship strength. For instance, if we observe that customer  $i$  gives items to customer  $j$ , it does not necessarily mean that  $i$  has a strong relationship with  $j$ .  $i$  could be generous (i.e.,  $i$  gives items to many people at the same time, not just  $j$ ), and/or  $j$  could be popular (i.e.,  $j$  receives items from many people other than  $i$  at the same time). We use generosity and popularity covariates to control for such effects. Also, if customers were active and diligent in the game, they would have more chances to interact and share items with other players. We use the log count of “monster fights”<sup>17</sup> to capture such a devotion effect. The event covariate  $Last\_2W$  is used to capture the change in customer interaction behaviors in weeks 24 and 25, the final two weeks before the server was shut down.

$X_{ijt}^{H[1]}$  is the vector of homophily covariates, which captures the effect of demographic and behavioral similarity between the two customers in the dyad. The nominal network covariates are calculated from customers' registration information (friendship and guild membership). The behavioral network covariates are calculated from customers' interaction

<sup>17</sup> “Monster fight” is gaming jargon that refers to the fight between players and virtual figures in the game.

records.

$\mathbf{X}_{ijt}^{P[1]}$  is the vector of cumulative past interactions.  $Cum\_Trade\_WK_{ijt}(\rho^S)$  captures customer  $i$ 's “inertia” of interaction with customer  $j$ .  $Cum\_Trade\_WK_{jit}(\rho_R^S)$  measures the impact of  $j$ 's “reciprocity” on  $i$ 's interaction decisions. Both covariates are calculated using equations (4) and (5). We estimate the model over a grid of  $\rho^S$  and  $\rho_R^S$  and choose the combination of  $\rho^S$  and  $\rho_R^S$  that yields the best model fit.

### Model 2 - Sharing Frequency Model

We model dyad  $\langle i, j \rangle$ 's sharing frequency in week  $t$  using a Poisson distribution with a log link function. In our specification, conditioning on the random effects  $\gamma_{\langle i, j \rangle} = a_i + b_j$ , the weekly count of sharing in dyad  $\langle i, j \rangle$  follows independent Poisson distribution:

$$f_{Y_{\langle i, j \rangle t} | \gamma_{\langle i, j \rangle}}(y_{\langle i, j \rangle t}^{SF} | \mathbf{X}_{\langle i, j \rangle t}^{[2]}, \boldsymbol{\beta}^{[2]}, \boldsymbol{\gamma}_{ij}^{[2]}) = \frac{e^{-\mu_{\langle i, j \rangle t} \times (\mu_{\langle i, j \rangle t})^{y_{\langle i, j \rangle t}^{SF}}}}{y_{\langle i, j \rangle t}^{SF}!}, \quad (1.8)$$

and the mean of the Poisson distribution is parameterized as

$$\mu_{\langle i, j \rangle t} = \exp \left( \underbrace{\mathbf{X}_{\langle i, j \rangle t}^{A[2]'} \cdot \boldsymbol{\beta}^{A[2]} + \mathbf{X}_{\langle i, j \rangle t}^{H[2]'} \cdot \boldsymbol{\beta}^{H[2]} + \mathbf{X}_{\langle i, j \rangle t}^{P[2]'} \cdot \boldsymbol{\beta}^{P[2]}}_{\text{fixed effects}} + \underbrace{a_i^{[2]} + b_j^{[2]}}_{\text{random effects}} + \underbrace{\epsilon_{ijt}^{[2]}}_{\text{error}} \right) \quad (1.9)$$

Table 1.4 displays all the covariates used in this model.  $\mathbf{X}_{\langle i, j \rangle t}^{A[2]}$  is almost the same as  $\mathbf{X}_{\langle i, j \rangle t}^{A[1]}$ . The only difference is that we replace dyad count covariates  $Dyad\_Count_{i,t}, \dots, Dyad\_Count_{j,t}$  with sharing count covariates  $Share\_Count_{i,t}, \dots, Share\_Count_{j,t}$ . The purpose remains the same: to capture the impact of “generosity” and “popularity” factors on customers' sharing frequency.

Compared with  $\mathbf{X}_{\langle i,j \rangle t}^{H[1]}$ , in  $\mathbf{X}_{\langle i,j \rangle t}^{H[2]}$  we replace  $Tenure\_Diff_{ij}$  and  $Tenure\_Avg_{ijt}$  with  $Tenure_{it}$  and  $Tenure_{jt}$ . We also add two new covariates:  $Level_{it}$  and  $Level_{jt}$ , which are customer  $i$  and  $j$ 's experience levels in week  $t$ . Moreover, we replace  $PK\_Count_{ijt}$  with a behavioral similarity covariate  $Mission\_Sim_{ijt}$ . It is defined as  $\frac{Mission\_Count_{ijt}}{Mission\_Count_{it} + Mission\_Count_{jt}}$ , where  $Mission\_Count_{it}$  and  $Mission\_Count_{jt}$  are the numbers of *unique* missions customer  $i$  and  $j$  joined in week  $t$ ;  $Mission\_Count_{ijt}$  is the number of *unique* missions  $i$  and  $j$  joined in together *at the same time* within week  $t$ . We use this covariate as a proxy of customers' interactions in the missions.<sup>18</sup>

Different from  $\mathbf{X}_{\langle i,j \rangle t}^{P[1]}$ , in  $\mathbf{X}_{\langle i,j \rangle t}^{P[2]}$  we use cumulative sharing count covariates  $Cum\_Share\_Count_{ijt}(\rho^{SF})$  and  $Cum\_Share\_Count_{jit}(\rho_R^{SF})$ , because they work well to predict weekly sharing frequency. They are calculated using equations (4) and (5). We determine the values of  $\rho^{SF}$  and  $\rho_R^{SF}$  following our methods for determining the values of  $\rho^S$  and  $\rho_R^S$  (i.e., grid search).

### Identification of Strong / Weak Ties

After estimating Models 1 and 2, we then use the estimated coefficients of homophily and interaction history parameters to calculate two tie strength measures:

$$\begin{aligned} RS_{\langle i,j \rangle t}^1 &= \exp \left( \mathbf{X}_{\langle i,j \rangle t}^{H[1]'} \cdot \hat{\boldsymbol{\beta}}^{H[1]} + \mathbf{X}_{\langle i,j \rangle t}^{P[1]'} \cdot \hat{\boldsymbol{\beta}}^{P[1]} \right) \\ RS_{\langle i,j \rangle t}^2 &= \exp \left( \mathbf{X}_{\langle i,j \rangle t}^{H[2]'} \cdot \hat{\boldsymbol{\beta}}^{H[2]} + \mathbf{X}_{\langle i,j \rangle t}^{P[2]'} \cdot \hat{\boldsymbol{\beta}}^{P[2]} \right), \end{aligned} \quad (1.10)$$

where  $\mathbf{X}_{\langle i,j \rangle t}^{H[m]}$  and  $\mathbf{X}_{\langle i,j \rangle t}^{P[m]}$  are the data vectors of Model  $m$ ;  $\hat{\boldsymbol{\beta}}^{H[m]}$  and  $\hat{\boldsymbol{\beta}}^{P[m]}$  are the vectors of the

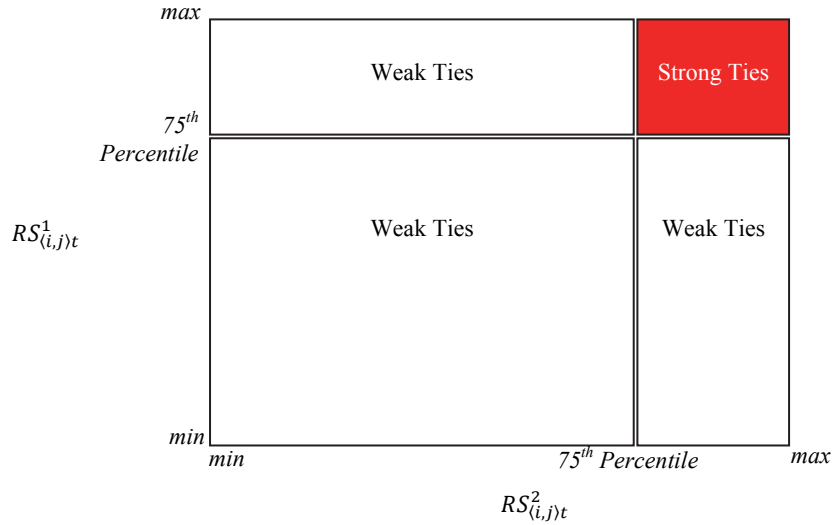
---

<sup>18</sup> In our data we can track whether two customers participated in the same mission at the same time and location. However, we still lack the direct evidence that these two customers were indeed interacting in the game.

estimated coefficients of Model  $m$ ,  $m = 1, 2$ .<sup>19</sup>  $RS^1_{\langle i,j \rangle t}$  measures dyad  $\langle i, j \rangle$ 's tie strength in terms of the customers' propensity to interact; and  $RS^2_{\langle i,j \rangle t}$  measures tie strength in terms of the customers' interaction frequency when they decide to interact.

After calculating tie strength measures, we use the following rule to identify the strong and weak ties (Figure 1.6): in each week we use the  $RS^1$  and  $RS^2$  measures to pick out the two sets of top 25 percentile<sup>20</sup> dyads. The dyads at the intersection of these two sets are marked as strong ties in that week.

**Figure 1.6**  
**The Identification of Strong/Weak Ties**



The rationale for this approach is that a dyad with a strong relationship is more likely to interact and at a higher frequency than a dyad with a weak relationship. If Models 1 and 2 are well specified and estimated,<sup>21</sup> the dyads labeled as strong ties should have a significantly higher likelihood of social interactions and higher interaction frequency than other dyads.

<sup>19</sup> We only include the significant coefficients in our calculation.

<sup>20</sup> We experimented with various cutoff values: top 10, 25, and 50 percentile. We found that the top 25 percentile value had the best distinguishing power.

<sup>21</sup> By “well specified and estimated” we mean the model should yield good goodness-of-fit according to criteria such as likelihood, AIC, and BIC.

After strong ties and weak ties are identified in each week and for each individual customer, we cluster each customer's active social ties into four cohesion subgroups (Figure 1.4). Here we only count the active ties. In other words, sharing behaviors must be observed in these ties. We then move on to generate various aggregate-level sharing and payment covariates of these four subgroups, which we then use in Model 3 and Model 4.

### Model 3 – Purchase Decision Model

We model customer  $i$ 's cash purchase decision with a logit model. We define customer  $i$ 's utility gained from her cash payment decision as

$$u_{it}^P = \underbrace{\mathbf{X}_{it}^{I[3]'} \cdot \boldsymbol{\beta}^{I[3]} + \mathbf{X}_{it}^{N[3]'} \cdot \boldsymbol{\beta}^{N[3]} + \mathbf{X}_{it}^{B[3]'} \cdot \boldsymbol{\beta}^{B[3]} + \mathbf{X}_{it}^{P[3]'} \cdot \boldsymbol{\beta}^{P[3]}}_{\text{fixed effects}} + \underbrace{a_i^{[3]}}_{\text{random effects}} + \underbrace{\epsilon_{it}^{[3]}}_{\text{error}}. \quad (1.11)$$

The purchase decision is

$$y_{it}^P = \begin{cases} 1, & u_{it}^P > 0 \\ 0, & \text{otherwise} . \end{cases}$$

Table 1.5 shows the covariates used in Model 3.

$\mathbf{X}_{it}^{I[3]}$  is the vector of individual characteristics. These covariates capture the impact of a customer's experience (duration of game play and level within the game) on her purchase decisions.  $\mathbf{X}_{it}^{N[3]}$  is the vector of nominal network characteristics. We use these covariates to analyze the impact of friends and guilds.  $\mathbf{X}_{it}^{B[3]}$  is the vector of behavioral network characteristics. These covariates will help us understand the influence of actively interacting customers. The vector of past payment decisions  $\mathbf{X}_{it}^{P[3]}$  has one covariate:  $Cum\_Paid\_WK_{it}(\rho^P)$ . It is calculated with equation (1.4) and (1.5), using the time-diminishing coefficient  $\rho^P$ . We estimate Model 3

on a grid of  $\rho^P$  and choose the  $\rho^P$  value that leads to the best model fit.

**Table 1.5**  
**Covariates Used in Models 3 and 4**

Covariates		Type	Model		Description
			3	4	
$X^I$	$Level_{it}$	Individual characteristics	x	x	Customer $i$ 's average level in week $t$
	$Mission\_Count_{it}$	Individual characteristics	x	x	The number of missions customer $i$ finishes in week $t$
	$Online\_Time_{it}(hr)$	Individual characteristics	x	x	The total time (hours) customer $i$ spends in the game in week $t$
	$Tenure_{it}$	Individual characteristics	x	x	Customer $i$ 's tenure up until week $t$
	$Tenure_{it}^2$	Individual characteristics	x	x	The squared term of customer $i$ 's tenure
	$Last\_2W$	Event	x	x	The indicator of week 24 and week 25
$X^N$	$Friend\_Count_{it}$	Nominal network	x	x	The number of friends customer $i$ has made up until week $t$
	$Guild\_Count\_L_{it}$	Nominal network	x	x	The number of guilds that customer $i$ leads in week $t$
	$Guild\_Member\_Count\_L_{it}$	Nominal network	x	x	The total size of the guilds that customer $i$ leads in week $t$
	$Guild\_Count\_M_{it}$	Nominal network	x	x	The number of guilds that $i$ has joined as a member up until week $t$
	$Guild\_Member\_Count\_M_{it}$	Nominal network	x	x	The total size of the guilds that $i$ has joined as a member up until week $t$
	$Share\_Dyad\_Count_{i,t}(ST)$	Behavioral network	x		The number of strong ties from customer $i$ to other customers
	$Share\_Dyad\_Count_{i,t}(WT)$	Behavioral network	x		The number of actively-sharing weak ties from customer $i$ to other customers in week $t$
	$Share\_Dyad\_Count_{i,t}(ST)$	Behavioral network	x		The number of active strong ties from other customers to customer $i$ in week $t$
	$Share\_Dyad\_Count_{i,t}(WT)$	Behavioral network	x		The number of active weak ties from other customers to $i$ in week $t$
	$Payment_{i,t}(ST)$	Behavioral network		x	The total amount of cash payments in week $t$ made by customers who have active strong ties from customer $i$ in week $t$
	$Payment_{i,t}(WT)$	Behavioral network		x	The total amount of cash payments in week $t$ made by customers who have active weak ties from customer $i$ in week $t$
	$Payment_{i,t}(ST)$	Behavioral network		x	The total amount of cash payments in week $t$ made by customers who have active strong ties toward customer $i$ in week $t$
	$Payment_{i,t}(WT)$	Behavioral network		x	The total amount of cash payments in week $t$ made by customers who have active weak ties toward customer $i$ in week $t$
$X^P$	$Cum\_Paid_{it}(\rho^P)$	Payment history	x		The cumulative number of weeks in which customer $i$ make cash payments during the past five weeks, calculated with time-diminishing parameter $\rho^P$
	$Cum\_Payment_{it}(\rho^{PA})$	Payment history		x	The cumulative payments made by customer $i$ during the past five weeks, calculated with time-diminishing parameter $\rho^{PA}$

#### Model 4 – Payment Amount Model

We model customer  $i$ 's weekly cash payment amount (if she pays) with a lognormal model. Conditioning on the random effect  $a_i$ ,

$$y_{it}^{PA} | X_{it}^{[4]}, \beta^{[4]}, \gamma_i^{[4]} \sim \text{i. i. d. lognormal}(\mu_{it}^{PA}, \sigma_{PA}^2), \quad (1.12)$$

and

$$\mu_{it}^{PA} = \underbrace{X_{it}^{I[4]'} \cdot \beta^{I[4]} + X_{it}^{N[4]'} \cdot \beta^{N[4]} + X_{it}^{B[4]'} \cdot \beta^{B[4]} + X_{it}^{P[4]'} \cdot \beta^{P[4]}}_{\text{fixed effects}} + \underbrace{a_i^{[4]}}_{\text{randome effects}} + \underbrace{\epsilon_{it}^{[4]}}_{\text{error}}. \quad (1.13)$$

Table 1.5 describes the covariates used in Model 4. The vector of individual characteristics,  $X_{it}^{I[4]}$ , is the same as  $X_{it}^{I[3]}$ ; and the vector of nominal network characteristics,  $X_{it}^{N[4]}$ , is the same as  $X_{it}^{N[3]}$ . In  $X_{it}^{B[4]}$ , the vector of behavioral network characteristics, we use the aggregate weekly payment made by the four cohesion subgroups to capture the group-to-one influence on customers' payment amounts. The cumulative past payment vector  $X_{it}^{P[4]}$  has one covariate:  $Cum\_Payment_{it}(\rho^{PA})$ . It is calculated with the time-diminishing coefficient  $\rho^{PA}$ . Similarly, we estimate Model 4 on a grid of  $\rho^{PA}$ , then choose the  $\rho^{PA}$  value that yields the best model fit.

### 1.5 Estimation

The three most commonly used methods to estimate generalized linear mixed models (GLMMs) are maximized likelihood (ML), penalized quasi-likelihood (PQL-1), and Markov chain Monte Carlo (MCMC) (Rabe-Hesketh and Skrondal 2010). We use SAS's GLIMMIX procedure with Laplace approximation (SAS 2011) to estimate all four models for three reasons. First, compared to other estimation methods (ML and MCMC), the approximation method used

in GLIMMIX is much faster. This is particularly important for us when we cross-validate our models, because we use a time-intensive algorithm. Second, compared to other approximation algorithms (e.g., pseudo-likelihood method or maximum likelihood with adaptive quadrature), although Laplace algorithm is slower to converge, it is more accurate and robust (Rabe-Hesketh and Skrondal 2010). Third, unlike pseudo-likelihood algorithms, the Laplace algorithm reports log likelihood and various goodness-of-fit measures such as AIC and BIC, which enable us to compare the models easily.

We estimate each of the four models (Models 1–4) in four steps (A–D). At each step we add in one type of parameter and compare the results (see Table 1.6).

**Table 1.6**  
**Model Estimation Results**

(1) Dyadic Level Models

<i>Model</i>	<i>Covariates</i>				<i>Goodness-of-Fit</i>		
	<i>Auxiliary</i>	<i>Nominal Network</i>	<i>Behavioral Network</i>	<i>Interaction History</i>	<i>-2Log Likelihood</i>	<i>AIC</i>	<i>BIC</i>
<i>1A</i>	x				10720.54	10740.54	10720.54
<i>Sharing Decision</i>	<i>1B</i>	x			10565.51	10609.51	10565.51
	<i>1C</i>	x		x	10381.91	10429.91	10381.91
	<i>1D</i>	x	x	x	10042.60	10094.60	10042.60
<i>2A</i>	x				31523.03	31543.03	31523.03
<i>Sharing Frequency</i>	<i>2B</i>	x			30126.13	30174.13	30126.13
	<i>2C</i>	x		x	28826.97	28878.97	28826.97
	<i>2D</i>	x	x	x	28668.74	28724.74	28668.74



**Table 1.6 (Continued)**

(2) Individual Level Models

<i>Model</i>	<i>Covariates</i>				<i>Goodness-of-Fit</i>		
	<i>Individual Characteristics</i>	<i>Nominal Network</i>	<i>Behavioral Network</i>	<i>Payment History</i>	<i>-2Log Likelihood</i>	<i>AIC</i>	<i>BIC</i>
<i>Payment Decision</i>	<i>3A</i>	x			1699.84	1715.84	1741.43
	<i>3B</i>	x	x		1683.88	1709.88	1751.46
	<i>3C</i>	x	x	x	1657.80	1691.80	1746.17
	<i>3D</i>	x	x	x	1636.22	1672.22	1729.79
<i>Payment Amount</i>	<i>4A</i>	x			1624.90	1642.90	1669.18
	<i>4B</i>	x	x		1625.17	1653.17	1694.05
	<i>4C</i>	x	x	x	1576.02	1612.02	1664.58
	<i>4D</i>	x	x	x	1560.13	1598.13	1653.61

At Step A we only consider the individual characteristics. At Step B we add the social network information obtained from customer registration data. At Step C we include the social network information calculated from customer social interaction data. Finally, at Step D we consider customer interaction and payment history. This stepwise approach enables us to check the power of different types of covariates to explain the variation in customer social interaction and purchasing behavior. Table 1.6 shows that the full models (Models 1D, 2D, 3D, 4D) yield the best fit according to log likelihood, AIC, and BIC measures. Therefore in subsequent subsections we only report the estimation results of full models. The full results of this section are available from the authors upon request.

#### *Model 1. Sharing Decisions*

Table 1.7 displays the results from Model 1. Generosity, popularity, and reciprocity have a significant impact on customers' sharing decisions.  $Dyad\_Count_i$  and  $Dyad\_Count_j$  are significant and positive. This tells us that if customer  $i$  were generous and/or customer  $j$  were

popular, then it is more likely that  $i$  would give items to  $j$ .  $Dyad\_Count_{.i}$  (customer  $i$ 's popularity) is slightly significant and negative. This can be explained by reciprocity: if  $i$  received items from many other customers, she would have reciprocated the favor and given other items to those customers. Therefore she would be less likely to give items to  $j$ .  $Dyad\_Count_{.j}$  ( $j$ 's generosity) is not significant in all four models. This is because usually  $i$  could not observe  $j$  giving items to others; therefore it had no impact on  $i$ 's sharing decisions.

**Table 1.7**  
**Sharing Decision Model (Model 1D) Estimation Results**

<i>Sharing Decision of <math>\langle i, j \rangle</math></i>	<i>Variable Type</i>	<i>Mean</i>	<i>STD</i>	
$Dyad\_Count_{.i}$	Generosity	0.04857	.015	***
$Dyad\_Count_{.j}$	Generosity	-0.01376	.015	
$Dyad\_Count_{.i}$	Popularity	-0.03126	.013	**
$Dyad\_Count_{.j}$	Popularity	0.03878	.013	***
$\log(Fight\_Count_i)$	Devotion	-0.00352	.011	
$\log(Fight\_Count_j)$	Devotion	0.06707	.011	***
$Last\_2W$	Event	0.886	.177	***
$Friend\_Count_i$	Nominal Network	-0.00913	.003	***
$Friend\_Count_j$	Nominal Network	-0.01028	.003	***
$Common\_Friend\_Count_{ij}$	Nominal Network	0.00093	.007	
$Guild\_Count_i$	Nominal Network	-0.03411	.022	
$Guild\_Count_j$	Nominal Network	-0.01222	.023	
$Common\_Guild\_Count_{ij}$	Nominal Network	0.05099	.030	*
$Guild\_Leader_i$	Nominal Network	-0.01658	.115	
$Guild\_Leader_j$	Nominal Network	-0.226	.122	*
$Dyad\_Duration_{ij}$	Behavioral Network	-0.5582	.033	***
$Dyad\_Duration_{ij}^2$	Behavioral Network	0.03381	.002	***
$Tenure\_Diff_{ij}$	Behavioral Network	-0.01635	.008	**
$Tenure\_Avg_{ij}$	Behavioral Network	0.03317	.011	***
$Online\_Time\_Overlap\_Pcnt_{ij}$	Behavioral Network	2.7665	.285	***
$PK\_Count_{ij}$	Behavioral Network	0.03003	.006	***
$Cum\_Trade\_WK_{ij}$ (Best model fit at $\rho^S = 0$ )	Interaction History	1.4495	.191	***
$Cum\_Trade\_WK_{ji}$ (Best model fit at $\rho_R^S = 0$ )	Interaction History	1.1404	.184	***

\*\*\*  $p < 0.01$ ; \*\*  $p < 0.05$ ; \*  $p < 0.10$

Second, devoted, actively interacting customers were more likely to share items with each other. We can tell this by the positive signs on the devotion covariate  $\ln(Fight\_Count_j)$

and on the behavioral similarity covariates  $Online\_Time\_Overlap\_Pcnt_{ij}$  and  $PK\_Count_{ij}$ .

This result is not surprising because actively interacting customers tended to spend more time together. They had more opportunities to see each other's belongings and exchange information and were thus more likely to share items.

Third, customers' propensity to share items decreased with the duration of the relationship and their tenures in the game. The signs and scales of  $Dyad\_Duration$  and  $Dyad\_Duration^2$  tell us that dyad duration has a negative impact on the probability of sharing, unless the relationship lasted a long time (longer than eight weeks). However, because the average dyad duration in the game was merely 3.24 weeks, most customers were less and less likely to share with the passage of time. We also find that an experience gap reduced customers' willingness to share ( $Tenure\_Diff$  has a negative sign). This result is not surprising because newcomers likely had few chances to interact with veterans. However, if both customers were experienced, they were more likely to share items ( $Tenure\_Avg$  has a positive sign). Interestingly,  $Last\_2W$  is significantly positive, which tells us that in the final two weeks of game play, customers were much more willing to share items. Customers knew that after the server shut down, all accounts would be closed and all items would be lost.

Finally, customers had a "long" memory about their past sharing decisions. Covariates  $Cum\_Share\_WK_{ij}(\rho^S)$  and  $Cum\_Share\_WK_{ji}(\rho_R^S)$  are both significantly positive. This means that if  $i$  and  $j$  shared items in the preceding five weeks, they were more likely to share again. We find that  $\rho^S = \rho_R^S = 0$  yields the best model fit. This means that to make her sharing decision,  $i$  would equally weigh her sharing interactions with  $j$  within each of the past five weeks. In other words, she had a "long" memory of past sharing decisions.

Model 2. Weekly Sharing Frequency

**Table 1.8**  
**Sharing Frequency Model (Model 2D) Estimation Results**

<i>Sharing Count of <math>\langle i, j \rangle</math></i>	<i>Variable Type</i>	<i>Mean</i>	<i>STD</i>	
<i>Share_Count<sub>i</sub></i>	Generosity	-0.00417	.0004	***
<i>Share_Count<sub>j</sub></i>	Generosity	-0.00165	.0004	***
<i>Share_Count<sub>i</sub></i>	Popularity	0.001821	.0003	***
<i>Share_Count<sub>j</sub></i>	Popularity	0.00051	.0003	*
$\log(\text{Fight\_Count}_i)$	Devotion	0.03386	.0046	***
$\log(\text{Fight\_Count}_j)$	Devotion	0.00454	.0045	
<i>Last_2W</i>	Event	0.1422	.0564	**
<i>Friend_Count<sub>i</sub></i>	Nominal Network	-0.00044	.0017	
<i>Friend_Count<sub>j</sub></i>	Nominal Network	0.004921	.0016	***
<i>Common_Friend_Count<sub>ij</sub></i>	Nominal Network	0.009633	.0029	***
<i>Guild_Count<sub>i</sub></i>	Nominal Network	-0.03502	.0147	**
<i>Guild_Count<sub>j</sub></i>	Nominal Network	-0.08926	.0144	***
<i>Common_Guild_Count<sub>ij</sub></i>	Nominal Network	0.08261	.0128	***
<i>Guild_Leader<sub>i</sub></i>	Nominal Network	0.02262	.0538	
<i>Guild_Leader<sub>j</sub></i>	Nominal Network	0.007043	.0586	
<i>Dyad_Duration<sub>ij</sub></i>	Behavioral Network	0.203	.0093	***
<i>Dyad_Duration<sub>ij</sub><sup>2</sup></i>	Behavioral Network	-0.00989	.0007	***
<i>Tenure<sub>i</sub></i>	Behavioral Network	-0.05563	.0086	***
<i>Tenure<sub>j</sub></i>	Behavioral Network	-0.02595	.0086	***
<i>Level<sub>i</sub></i>	Behavioral Network	-0.00271	.0009	***
<i>Level<sub>j</sub></i>	Behavioral Network	0.001316	.0008	
<i>Online_Time_Overlap_Pcnt<sub>ij</sub></i>	Behavioral Network	3.1088	.0879	***
<i>Mission_Similarity<sub>ij</sub></i>	Behavioral Network	0.5907	.0799	***
<i>Cum_Share_Count<sub>ij</sub></i> (Best model fit at $\rho^{SF} = 0$ )	Interaction History	0.006194	.0008	***
<i>Cum_Share_Count<sub>ji</sub></i> (Best model fit at $\rho_R^{SF} = 0$ )	Interaction History	0.008347	.0010	***

\*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.10$

Most Model 2 results mirror those of Model 1. The two popularity covariates *Share\_Count<sub>j</sub>* and *Share\_Count<sub>i</sub>* are significant and positive, which means the more frequently customer *i* and customer *j* received items from other customers, the more frequently they would share. The two generosity covariates, *Share\_Count<sub>i</sub>* and *Share\_Count<sub>j</sub>*, are significant and negative, which again reveals a “diluting” effect: if customer *i* and *j* shared with other customers very often, it would reduce the sharing frequency between them. Customer *i*’s

“devotion” covariate  $\ln(Fight\_Count_i)$  and behavioral similarity covariates  $Online\_Time\_Overlap\_Pcnt_{ij}$  and  $Mission\_Similarity_{ij}$  are all significantly positive. This means that if customers were active and spent a lot of time together, they would share items more often. We also find that over time customers shared less and less often (negative signs for  $Tenure_i$ ,  $Tenure_j$ , and  $Level_i$ ). One possible explanation is that when customers advanced in the game, their belongings became more and more valuable; therefore they would not share as often as they did in the early stages.<sup>22</sup>  $Last\_2W$  is significantly positive. This means that people were sharing items more often during the final two weeks, because they knew that they would lose their items when the server shut down.  $Cum\_Share\_Count_{ij}$  and  $Cum\_Share\_Count_{ji}$  are both significantly positive, and we get the best model fit at  $\rho^{SF} = \rho_R^{SF} = 0$ . Therefore, in terms of sharing frequency, customers had a “long” memory about past sharing experiences.

However, two things catch our attention in Model 2. First, the friendship and guild membership covariates in the sharing frequency model become more significant than their counterparts in Model 1.  $Common\_Friend\_Count_{ij}$  and  $Common\_Guild\_Count_{ij}$  are positive and significant, which means that if two customers shared many of the same friends and joined many guilds together, they would share items more frequently.  $Guild\_Count_i$  and  $Guild\_Count_j$  have negative signs, which reveals the “diluting” effect again. The positive sign of  $Friend\_Count_j$  tells us that if a customer was popular and had many friends, other customers shared with her more frequently.

One explanation for this discrepancy is the existence of two different types of sharing behaviors: “casual” sharing and “serious” sharing. When customers met in the game, they might share (probably low-value) items casually, as a means of socializing. Therefore, friendship and

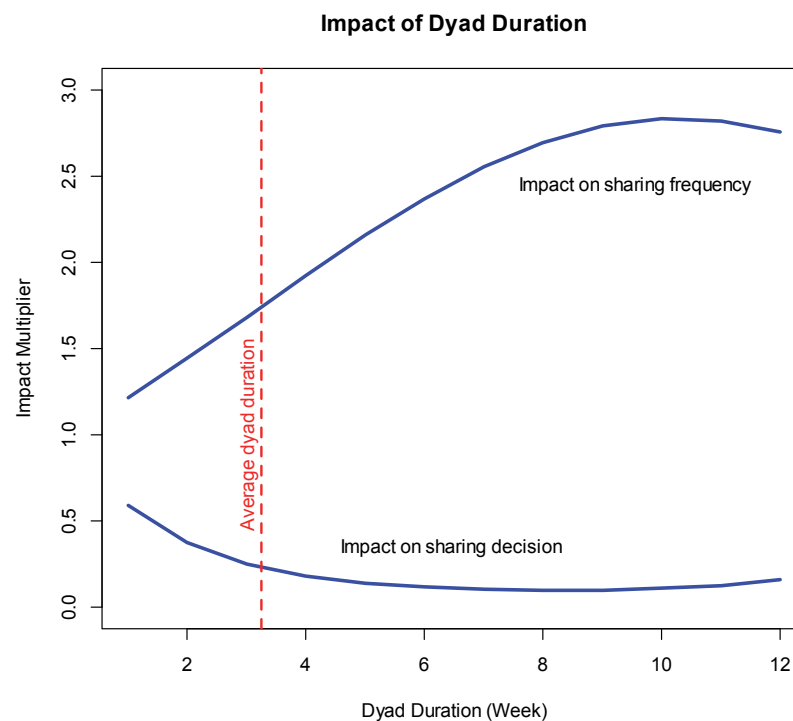
---

<sup>22</sup> Unfortunately we do not have complete records about the value of items shared each time. This is one limitation of our research.

guild membership might have had no significant impact on their decision to share. However, among customers who did share, high weekly sharing frequency could indicate a serious relationship. In such a case, social connections become important. Therefore, social connection covariates become more significant in the sharing frequency model.

Second, the signs of *Dyad\_Duration* and *Dyad\_Duration*<sup>2</sup> in the sharing frequency model (Model 2) are different from those in the sharing decision model (Model 1). Model 1 tells us that over time, customers were less and less likely to share items with each other. Model 2, however, tells us that customers with enduring relationships shared more often in each week if they decided to share, with sharing activities increasing each week until week 10 (Figure 1.7).

**Figure 1.7**  
**The Impact of Dyad Duration on Customers' Sharing Decision and Frequency**



This result, again, can be explained by “casual” versus “serious” sharing behaviors. With the passage of time, more and more customers left the game. Thus, the proportion of casual sharing likely decreased among the remaining customers. Those who remained in the game tended to

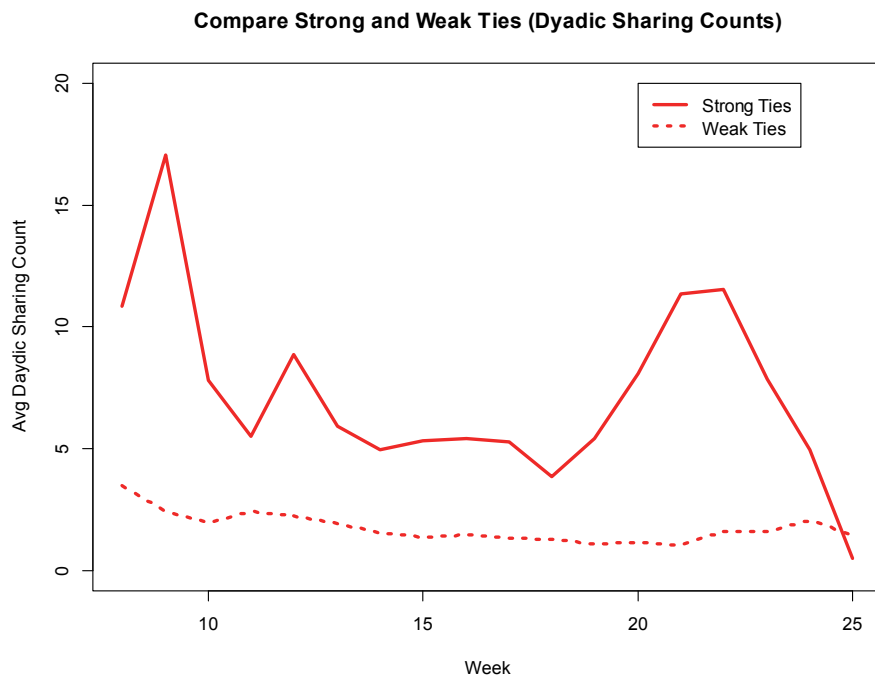
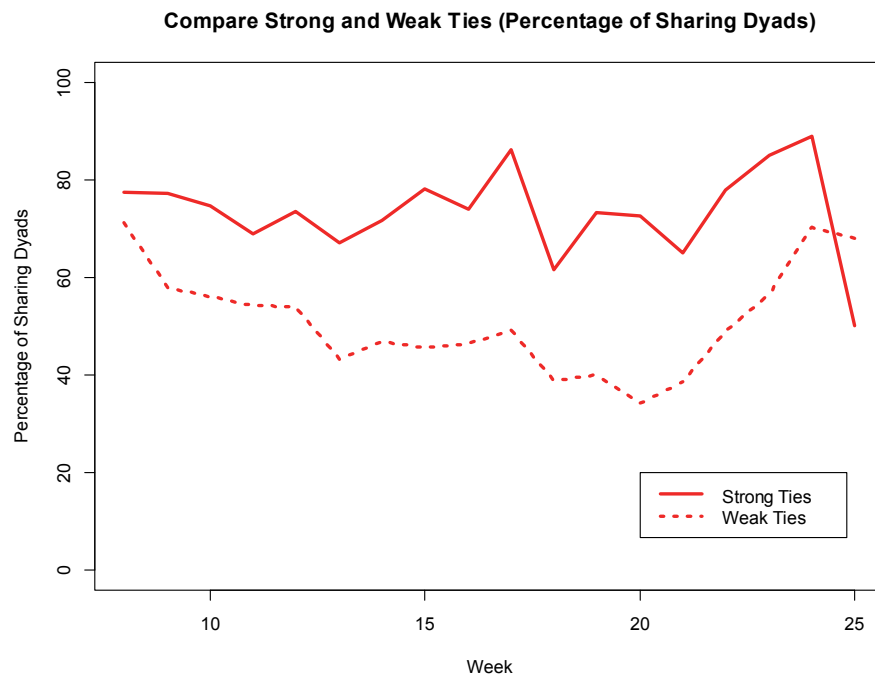
cultivate strong relationships and were more likely to share frequently within each week.

#### *Strong Tie and Weak Tie Identification*

After estimating Models 1 and 2, we use equation (1.10) to calculate  $RS_{\langle i,j \rangle,t}^1$  and  $RS_{\langle i,j \rangle,t}^2$ , the two tie strength measures of dyad  $\langle i,j \rangle$  in week  $t$ . We can then identify all the strong and weak ties with these two measures. Note that these measures are both directional. Therefore dyad  $\langle i,j \rangle$  and  $\langle j,i \rangle$  might have different tie strength at the same time. This leads to an interesting question: is relationship strength, in general, symmetric? We find that most dyads have symmetric tie strength. Among all dyad-week observations, approximately 26% of cases feature mutually strong relationships; about 63% feature mutually weak relationships. In only 11% cases, dyads have asymmetric relationship strength. This demonstrates that reciprocity is an important feature in online relationships.

In Figure 1.8 we compare the percentage of sharing dyads and the average weekly sharing frequency among customers with strong ties and weak ties. We also run a t-test to compare the means of these two groups. From the graphs and the t-test results, we see that the customers with strong ties were more likely to share and tended to share more often. These results demonstrate the effectiveness of using the two tie strength measures developed in our research.

**Figure 1.8**  
**Comparison of Strong and Weak Ties in Sharing Behaviors**



T-test shows that strong ties have significantly higher percentage of sharing dyads and higher sharing frequency than weak ties ( $p < 0.0001$ ).



### Model 3. Cash Payment Decisions

**Table 1.9**  
**Cash Payment Decision Model (Model 3D) Estimation Results**

<i>Payment Decision</i>	<i>Variable Type</i>	<i>Mean</i>	<i>STD</i>	
<i>Level</i>	Experience	0.006	.005	
<i>Mission_Count</i>	Devotion	0.003595	.001	**
<i>Online_Time(hr)</i>	Devotion	0.01198	.002	***
<i>Tenure</i>	Experience	-0.1399	.048	***
<i>Tenure</i> <sup>2</sup>	Experience	0.002297	.002	
<i>Last_2W</i>	Event	-0.9443	.543	*
<i>Friend_Count</i>	Nominal Network	0.001815	.007	
<i>Guild_Count_L</i>	Nominal Network	0.2508	.498	
<i>Guild_Member_Count_L</i>	Nominal Network	-0.00785	.010	
<i>Guild_Count_M</i>	Nominal Network	0.06774	.100	
<i>Guild_Member_Count_M</i>	Nominal Network	0.001088	.002	
<i>Share_Dyad_Count<sub>i</sub>(ST)</i>	Behavioral Network	0.3061	.133	**
<i>Share_Dyad_Count<sub>i</sub>(WT)</i>	Behavioral Network	0.09156	.047	*
<i>Share_Dyad_Count<sub>i</sub>(ST)</i>	Behavioral Network	-0.1723	.122	
<i>Share_Dyad_Count<sub>i</sub>(WT)</i>	Behavioral Network	0.08004	.047	*
<i>Cum_Paid_WK</i> (Best model fit at $\rho^P = 5$ )	Payment History	0.8109	.164	***

\*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.10$

Model 3 produces three major findings. First, devoted customers were more likely to pay cash. *Mission\_Count* and *Online\_Time(hr)* are both significantly positive. This tells us that if customers spent more time in the game and participated in more missions, they were more willing to purchase items with cash.

Second, an individual's social ties might be exerting some group-to-one influence on that customer. Recall that *Share\_Dyad\_Count<sub>i</sub>(ST)*, *Share\_Dyad\_Count<sub>i</sub>(WT)*, and *Share\_Dyad\_Count<sub>i</sub>(WT)* are the numbers of customers who had active connections with customer *i* (with various direction and tie strength). These three covariates are all positive. This tells us that a customer was more likely to purchase with cash if she was actively sharing with others.

Third, no friendship and guild membership covariates (i.e., nominal network

characteristics) are significant. This reminds us that when modeling customer behaviors in a social network, we cannot simply use the covariates generated from registration information. Activity-based covariates must also be included in the model.

Fourth, we find that customers became less willing to make cash payments over time in the game. *Tenure* is significant and negative. This makes sense because with time, the game became harder and items became more costly. Some customers might have quit and left the game. Those who chose to stay might have become less ambitious to get to higher levels, which required larger cash investment in items. They might have continued to play simply to have fun with their friends. As a result, customer willingness to purchase decreased over time. *Last\_2W* is slightly significant and negative. Apparently when customers knew that the server would soon shut down, they were unwilling to invest in items using real cash.

Finally, the cumulative payment decision covariate *Cum\_Paid\_WK* calculated with  $\rho^P = 5.0$  is significantly positive and yields the best model fit. This tells us two things: (1) a customer's past purchase decisions had significant impact on her later purchasing decisions, and (2) customers tended to have a "short" memory regarding their past payment decisions ( $\rho^P = 5.0$  means that almost all weight is put on the most recent purchasing decision).

#### *Model 4. Cash Payment Amount*

Table 1.10 shows that *Level* is significant and negative. Again this indicates that customers became less willing to pay cash for items when the game got harder and items got more expensive. *Online\_Time(hr)* is significant and positive; therefore "devotees" tended to pay more cash to purchase items. Interestingly, we find that *Last\_2W* is significantly positive. It seems that although in the final two weeks customers, on average, became less willing to pay cash, those who did pay real money paid more. Possibly some customers wanted to advance

faster and get to the end of the game before the server shut down; hence, they were willing to pay more. The cumulative past cash payment, *Cum\_Payment* calculated with  $\rho^{PA} = 5.0$ , is significantly positive and yields the best model fit. This tells us that customers had “short” memory when they decided how much real money to spend on virtual items.

**Table 1.10**  
**Cash Payment Amount Model (Model 4D) Estimation Results**

<i>Payment</i>	<i>Variable Type</i>	<i>Mean</i>	<i>STD</i>	
<i>Level</i>	Experience	-0.01683	.004	***
<i>Mission_Count</i>	Devotion	0.001411	.001	
<i>Online_Time(hr)</i>	Devotion	0.003478	.002	**
<i>Tenure</i>	Experience	0.017	.042	
<i>Tenure<sup>2</sup></i>	Experience	-0.00231	.002	
<i>Last_2W</i>	Event	1.2748	.596	**
<i>Friend_Count</i>	Nominal Network	-0.00311	.004	
<i>Guild_Count_L</i>	Nominal Network	0.1007	.323	
<i>Guild_Member_Count_L</i>	Nominal Network	-0.0005	.006	
<i>Guild_Count_M</i>	Nominal Network	0.03287	.082	
<i>Guild_Member_Count_M</i>	Nominal Network	-0.00048	.002	
<i>Payment<sub>i</sub>.(ST), 000</i>	Behavioral Network	-0.236	.104	**
<i>Payment<sub>i</sub>.(WT), 000</i>	Behavioral Network	-0.3706	.066	***
<i>Payment<sub>i</sub>.(ST), 000</i>	Behavioral Network	0.2929	.097	***
<i>Payment<sub>i</sub>.(WT), 000</i>	Behavioral Network	0.4473	.063	***
<i>Cum_Payment</i> (Best model fit at $\rho^{PA} = 5$ )	Payment History	0.000214	.00003	***

\*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.10$

The key components of Model 4 are the four behavioral network covariates:

*Payment<sub>i</sub>.(ST)*, *Payment<sub>i</sub>.(WT)*, *Payment<sub>i</sub>.(ST)*, and *Payment<sub>i</sub>.(WT)*. We use them to capture the group-to-one social influence on a customer’s purchasing behaviors. Model 4 produces several interesting results.

First, the group-to-one social influence pattern is complex in the context of virtual product purchasing, especially when sharing is allowed. The covariates *Payment<sub>i</sub>.(ST)* and *Payment<sub>i</sub>.(WT)* are both negative. This result indicates that if customer *i* saw other players to whom she gave items spending real cash to purchase items, she would spend less cash. This

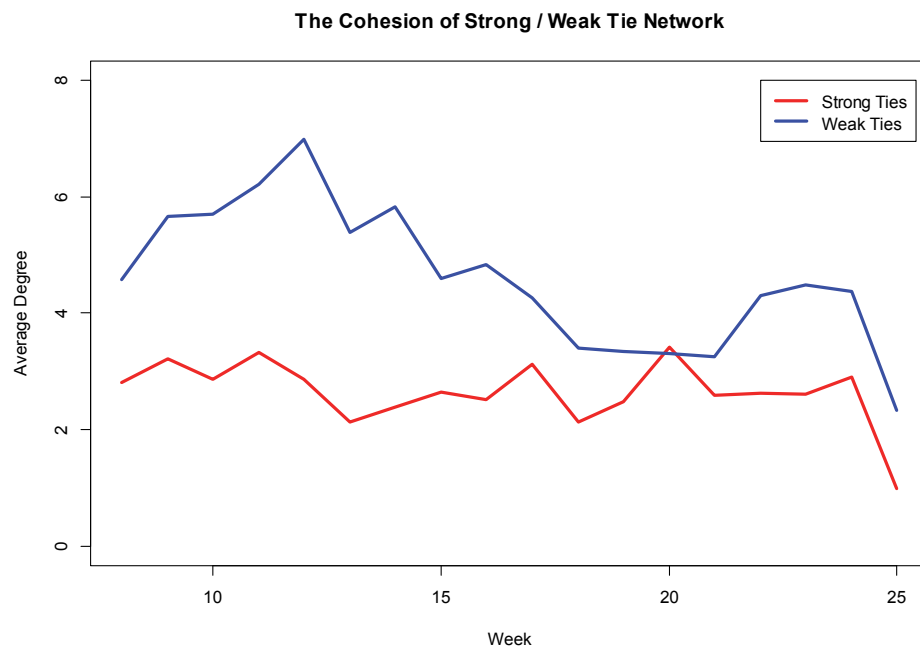
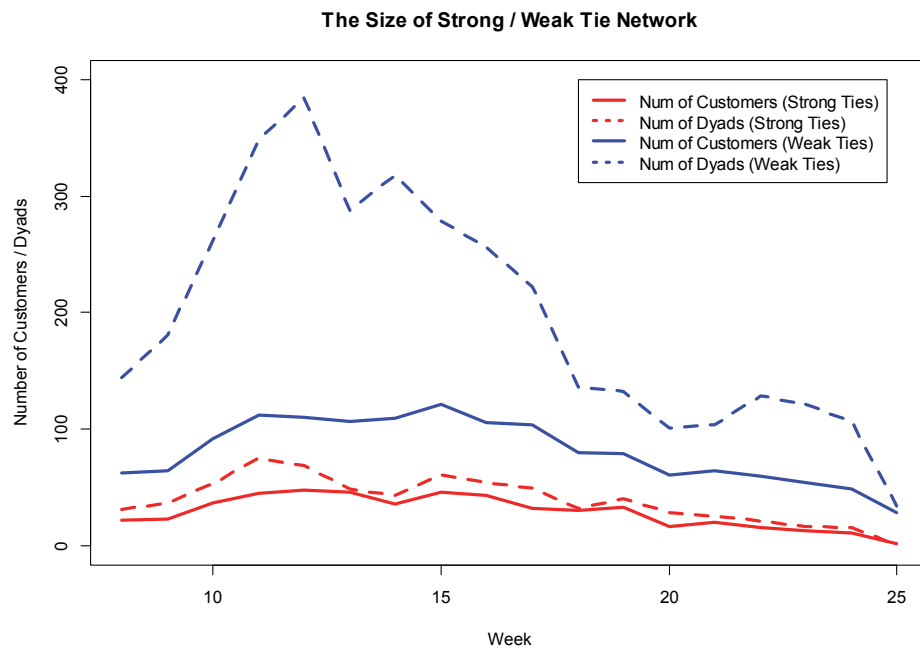
seems counterintuitive at first glance, because past research (e.g., Nair, Manchanda, and Bhati, 2010; Trusov, Bodapati, and Bucklin, 2010) tells us that customers who have influence on each other should have similar purchasing behaviors (i.e., the coefficients should be positive).

However, if we take into account the reciprocity factor in social relationships, this phenomenon makes sense. When a customer gives items to others, she expects them to reciprocate the favor and give her items that she needs in return. With such an expectation, she might reduce her cash expenditures if she sees that the customers who receive items from her are purchasing items with cash.  $Payment_i(ST)$  and  $Payment_i(WT)$  are both positive, which means that if customer  $i$  received items from customers who made cash purchases, she would pay more cash to purchase items as well. Again, reciprocity seems to be at play. If other customers are buying and giving items to customer  $i$ , she would be motivated to spend more on items so she can return the favor.

Interestingly,  $Payment_i(ST)$  (+) and  $Payment_i(WT)$  (+) have higher absolute values than  $Payment_i(ST)$  (−) and  $t_i(WT)$  (−), respectively. Hence, if customers gave and received items in the same week, the net impact on their payment amount could be positive.

Second, weak ties, as a whole, are more influential than strong ties. From the scales of the four covariates, we see that the weak tie subgroups have stronger influence than the strong tie subgroups. This, again, looks counterintuitive at first glance, because strong ties are supposed to carry greater influence than weak ties. This could be explained by the difference in the size and the cohesiveness of the strong and weak tie subgroups. Figure 1.9 plots the size and the average degree (as a measure of group cohesiveness) of the entire strong tie network and weak tie network in each week. It is quite obvious that the weak tie network is more widespread and more cohesive than the strong tie network.

**Figure 1.9**  
**Comparison of the Sizes of Strong Tie Network and Weak Tie Network**



In the strong tie–weak tie literature (e.g., Granovetter 1973; Bakshy et al. 2012), researchers have shown that weak ties play an important role in the dissemination of information. This is particularly true in our case: customers had more opportunities to see various items and exchange information through their weak ties than through their strong ties. As a result, weak ties, as a whole, had a stronger impact on customers’ purchase behaviors than strong ties.

Finally, we find that only actively interacting customers (customers who share items in that week) had a significant influence on each other’s purchasing behaviors. As a test, we include both active and inactive ties into the calculation of the four subgroup payment covariates. None of the four revised covariates is significant.

#### *Managerial Implications for Gaming Companies*

Based on our findings, we recommend five actions that gaming companies can take to increase their revenues. First, companies should encourage their customers to stay longer and be more active in the games, because this will directly increase customer propensity to pay and the amount they pay. Second, companies should encourage their customers to interact with as many other customers as possible, including casual interactions. Interactions enable customers to see new products and exchange information, which stimulates purchases. Third, they should encourage customers to share items with each other, because the net impact of sharing activities on revenues is positive. Fourth, gaming companies should foster customers’ initial interactions. After establishing relationships, customers will continue on their own, because they have a long memory of past interactions. Finally, following a customer cash purchase, companies should provide monetary incentives (e.g., coupon with a short expiration period) to the customer to spur additional purchases. Our research shows that customers have a short memory of past purchases, meaning that only the most recent purchase has an impact on their next purchase. Thus,

companies should offer incentives to encourage continuous spending.

## **1.6 Robustness Tests, Cross-Validation, Causality Inference, and Policy Simulations<sup>23</sup>**

### *Robustness Tests*

#### (1) Tests on the left truncation

We run a series of robustness tests to assess the impact of left truncation on our estimation results. In these tests we estimate Models 1D–4D (all full models) on three datasets: data of truncated customers/dyads, data of non-truncated customers/dyads, and data of all customers/dyads.

We find that some of the coefficients (mainly friend and dyad counts) estimated on truncated customer data become insignificant (with  $p$ -values above 0.10). This might be due to the small number of observations of truncated customers. However, we do not observe any sign changes (significant but with different signs). Therefore it seems that left truncation has only a small impact on our estimation results.

#### (2) Tests on lagged payments

We test a different specification of Model 3D and Model 4D by including the lagged subgroup payment covariates in the models. The coefficients of lagged covariates are all insignificant. Hence, a customer's current purchasing behaviors are not significantly influenced by other customers' previous purchasing behaviors.

### *Cross-Validation*

It is difficult to employ a usual holdout set approach to validate our models because we have a small sample across a short observation period. If we randomly assigned customers to an estimation set and a holdout set, the sample size of each group would be very small, not to

---

<sup>23</sup> Detailed results of this section are available from the authors upon request.

mention the difficulties of drawing representative random samples from an interconnected network. Using the observations in the final  $T$  weeks as our holdout set also creates difficulties. The server was shut down at the end of our observation period, and customers knew about it. This awareness that their accounts would be closed within a few weeks might have caused customers to behave differently.

Taking these issues into account, we use a “leave-one-out” (LOO) approach to cross-validate our models. First, we loop over all the dyads in our dataset to validate our sharing decision and frequency models (Models 1D and 2D). Each loop consists of three steps: (1) take out one dyad and estimate the models with the data of all the remaining dyads; (2) use estimated coefficients to predict the left-out dyad’s probability to share and the sharing frequency if they decided to share; and (3) compare the predicted values with the observed values. The loop continues until we finish all the dyads in our sample. Second, we use the same LOO approach to validate the payment decision and amount models (Models 3D and 4D). Similarly, at each loop we leave out one individual customer, then do the estimation and prediction, and compare the predicted values and the observed values for the deleted customer (See Table 1.11).<sup>24</sup>

The validation results show that with our models a company can identify interacting dyads and paying customer with much higher accuracy than a “model-free” approach in which the company randomly chooses a number of customers. For instance, Model 1D correctly identifies 72.61% of the interacting dyads. Using Model 2D, a company can detect 37.56% of the most actively interacting dyads (top 25 percentile). In terms of payment prediction, our models perform even more impressively. With Model 3D, a company can successfully pinpoint 53.22% of the payers in each week. Model 4D can help capture 51.35% of the most valuable customers

---

<sup>24</sup> Note that LOO cross-validation can be time-consuming. We ran our program on a Linux server, and it took approximately 5.5 days of CPU time to validate all four models.



(top 25 percentile in cash payments).

**Table 1.11**  
**Results of Cross-Validation**

(1) Sharing Decision Model

Model	% of Correct Identification of Interacting Dyads
Model-Free	51.99
Model 1A	62.96
Model 1B	65.18
Model 1C	65.39
Model 1D	72.61

(2) Sharing Frequency Model

Model	MSE	% of Successful Identification of Actively-Sharing Dyads	
		Top 50 Percentile	Top 25 Percentile
Model-Free	143.92	50.00	25.00
Model 2A	151.80	58.11	35.14
Model 2B	142.68	58.11	22.09
Model 2C	130.05	58.56	45.95
Model 2D	128.55	57.66	51.35

(3) Payment Decision Model

Model	% of Correct Identification of Payers
Model-Free	23.56
Model 3A	40.80
Model 3B	46.78
Model 3C	47.32
Model 3D	53.22

(4) Payment Amount Model

Model	MSE	% of Successful Identification of High-Value Customers	
		Top 50 Percentile	Top 25 Percentile
Model-Free	1,361,452	50.00	25.00
Model 4A	1,319,648	58.11	35.14
Model 4B	1,326,134	58.11	22.09
Model 4C	906,845	58.56	45.95
Model 4D	883,128	57.66	51.35

## *Causality Inference*

Model 4D shows a strong correlation between an individual customer's payments and the payments made by the customers actively interacting with her. However, further investigation is needed to ensure that such a correlation is caused only by the social influence (peer effects) and not by other possible sources of correlation, such as endogenous group formation, correlated unobservables, or simultaneity (Moffitt 2001).

### (1) Endogenous group formation

Customers with similar characteristics might behave in a similar manner. Therefore, the correlation between payments might actually reflect customers' similarity (homophily) instead of social influence. This, however, is unlikely in our case. First, we note that the coefficients of  $Payment_{i,}(ST)$  and  $Payment_{i,}(WT)$  are negative, and the coefficients of  $Payment_{i,}(ST)$  and  $Payment_{i,}(WT)$  are positive, which cannot be explained by homophily effects. Moreover, in Model 4D we have already included the customer random effect  $a_i$ , which could help explain away the possible correlation due to unobserved homophily factors. Therefore, endogenous group formation seems unlikely to be the cause of the correlation in customers' payments.

### (2) Correlated unobservables

If customers' payments are driven by some unobservable factors (e.g., promotion or seasonality), their payments could be correlated. We rule out this possibility for the following four reasons: First, the game analyzed in our research is small and promotion-free. Only a small gift card of game money was provided to all newcomers when they started. Second, we are analyzing the peer effects from four different cohesion subgroups on one individual customer. It is unlikely that all the customers in these four subgroups are correlated with the same unobservables at the same time—not to mention that the members of these groups vary with time.

Third, two subgroup payment coefficients are positive and two are negative. Evidently, correlated unobservables are not the cause of correlated payments. Finally, we re-estimate Model 4D with weekly fixed effects, but none of the fixed effects are significant ( $p \leq 0.10$ ). This means that temporal factors such as promotions or seasonality do not have an impact on customers' payments. Thus, we rule out correlated unobservables as a possible source of the correlation in customer payment amounts.

### (3) Simultaneity

If an individual customer's payments and the payments made by her four cohesion subgroups are contemporaneously interdependent, the subgroup payment coefficients could be biased upwards (Nair, Manchanda, and Bhatia 2010). To investigate whether contemporaneous interdependence exists in our data, we adopt a system-of-equations approach. For example, if we want to check whether covariate  $Payment_{i,t}(ST)$  is endogenous, we estimate the following two equations together:

$$\begin{cases} \ln(Payment_{it}) = \mathbf{X}_{it}' \cdot \boldsymbol{\beta}_1 + \gamma_1 \cdot Payment_{i,t}(ST) + a_i + \epsilon_{it,1} \\ Payment_{i,t}(ST) = \mathbf{X}_{it}' \cdot \boldsymbol{\beta}_2 + a_i + \epsilon_{it,2} \end{cases}, \quad (1.14)$$

where  $\mathbf{X}_{it}$  is the data vector of customer  $i$  in week  $t$ , *excluding* the four cohesion subgroup payment covariates;  $\boldsymbol{\beta}_1$  and  $\boldsymbol{\beta}_2$  are the coefficient vectors;  $a_i$  is customer  $i$ 's random effect; and  $\epsilon_{it,1}$  and  $\epsilon_{it,2}$  are the error terms. If  $\epsilon_{i1,t}$  and  $\epsilon_{i2,t}$  are not significantly correlated, we will have evidence that  $\ln(Payment_{it})$  and  $Payment_{i,t}(ST)$  are not contemporaneously interdependent. We run this test on all four subgroup payment covariates. (See Table 1.12.)<sup>25</sup>

---

<sup>25</sup> We do not use five equations to test all four subgroup payment covariates altogether because we do not have enough observations to do so.

**Table 1.12**  
**Endogeneity Tests on Subgroup Payment Covariates**

Test	Endogenous Relationship Tested	Correlation between $\epsilon_{i1,t}$ and $\epsilon_{i2,t}$	
		Coefficient	$Prob >  r $
1	$Payment_i$ and $Payment_{i.(ST)}$	-0.00055	0.9906
2	$Payment_i$ and $Payment_{i.(WT)}$	0.00018	0.9970
3	$Payment_i$ and $Payment_{.i}(ST)$	-0.00690	0.8838
4	$Payment_i$ and $Payment_{.i}(WT)$	0.06365	0.1772

We find that in all four tests  $\epsilon_{i1,t}$  and  $\epsilon_{i2,t}$  are not significantly correlated, which means that simultaneity is not a serious issue in this research. This finding coincides with our life experience: group-to-one influence is typically much stronger than one-to-group influence.

#### *Policy Simulation*

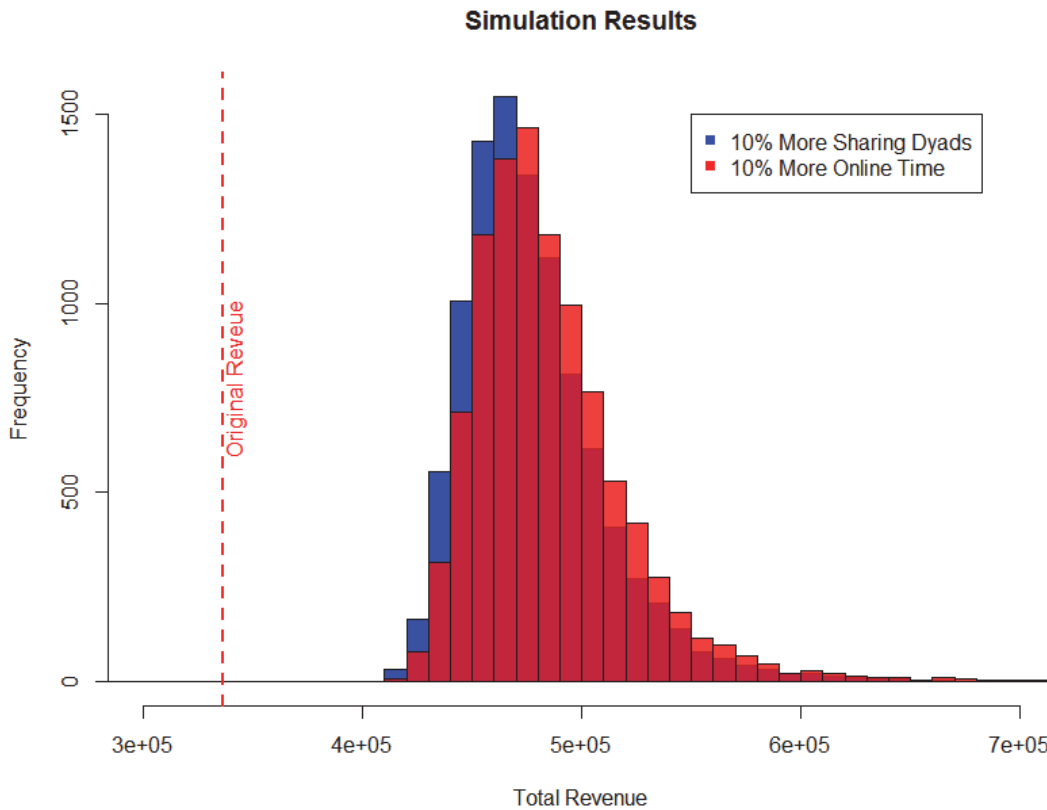
We design two policy simulations to demonstrate our framework's usefulness for increasing gaming company revenues. We suppose that a company is considering two separate strategies: (1) making customers more actively interact with each other (e.g., making it easier for customers to share their items) or (2) making customers spend more time in the game (e.g., using a customer loyalty program or increasing the chance to get windfall prizes in the game).

Accordingly, in the first simulation we assume that the company increases the number of actively interacting dyads by 10%. In the second simulation, we assume that the company induces its customers to spend 10% more time in the game. To simplify the operation, we make two assumptions. First, the newly added payment cases all occur at the end of each week; therefore they have no impact on previous purchases. Second, the distributions of newly-added payment amounts have the same variance. This variance is calculated from the data.

The simulation results show that increasing actively interacting dyads by 10% could lift revenue by 42.25%, and increasing customers' online time by 10% could lead to a 44.56%

revenue increase (Figure 1.10). Thus, increasing customers' online time is slightly more effective. The covariate *Online\_Time(hr)* is significant in both Model 3D and Model 4D, indicating that customers' online time not only boosts customers' propensity to pay, but also raises their payment amounts if they do pay. Such a difference is quite obvious when we compare the total revenues from these two strategies.

**Figure 1.10**  
**Policy Simulation Results**



## 1.7 Discussion

This study makes three primary contributions to the current social network literature. First, to the best of our knowledge, our research is the first to reveal the group-to-one social influence of strong and weak ties in the context of virtual product purchasing. We demonstrate the complexity of social influence patterns in such a setting, especially when sharing is allowed.

These findings could also shed light on customer purchasing behaviors of tangible products in the physical world in a social network context.

Second, we develop a comprehensive framework to combine social interaction models, social influence models, and tie strength measures. This framework enables us to see a more complete picture of customer connections, interactions, and influence within a social network. This framework de-composes the revenue generation process, thus enabling a company to leverage the power of social networks to increase revenues, as illustrated in our simulations.

Finally, we propose a tie strength measure based on social connection and interaction information and show that it performs well in distinguishing strong ties from weak ties. Our tie strength measure, which is directional, continuous, and time-varying, contains rich information about the characteristics of customer relationships. This measure is also modeled as the function of various factors. These findings can help companies cultivate enduring and profitable relationships among their customers.

Our research, however, is not without limitations. First, even though we have shown that the influence of left truncation is small, our estimation results might still be biased. Second, we only use the counts of social interactions (sharing of items) to measure customer tie strength, which might only provide a partial picture. Because of a lack of data on item value, we are unable to consider the values of items exchanged through those interactions. Finally, we could not directly observe customers' interactions in their game missions. We also do not have information about their communications inside and outside the game.<sup>26</sup>

Our framework could be extended in the following three directions. First, we could incorporate a latent state model, such as a hidden Markov model (HMM), into our framework.

---

<sup>26</sup> Modern game players use free online audio/video communication services (such as MSN or Skype) to coordinate their actions in the games. Only a few players still use texting within the game. Even so, most video game companies choose not to record such text communications due to privacy protection concerns.

With such a modification, our approach would enable a better understanding of the dynamic nature of customer interactions and influences in a social network. Second, our research could be extended to investigate how customers influence each other within a triadic structure. This extension, however, might prove to be highly challenging. Finally, we could estimate all four models jointly, which would enable us to directly model the relationship between customers' social interactions and purchases.

## REFERENCES

- Ansari, Asim, Oded Koenigsberg, and Florian Stahl (2011), "Modeling Multiple Relationships in Social Networks," *Journal of Marketing Research*, 48(August), 713–28.
- Austin, Scott (2011), "Facebook, Groupon, Zynga: Off-The-Chart Revenue," *Venture Capital Dispatch*, WSJ Blogs, online source: <http://blogs.wsj.com/venturecapital/2011/02/26/facebook-groupon-zynga-off-the-chart-revenue/>
- Bakshy, Eytan, Itamar Rosenn<sup>1</sup>, Cameron Marlow, and Lada Adamic (2012), "The Role of Social Networks in Information Diffusion," In *Proceedings of the 21st ACM conference on the World Wide Web*. ACM.
- Bartle, Richard (2003), *Designing Virtual Worlds*, New Riders Games.
- Chang, Victoria and Haim Mendelson (2010), "Social Games," Stanford Graduate School of Business Case.
- Christakis, Nicholas A. and James H. Fowler (2007), "The Spread of Obesity in a Large Social Network over 32 Years," *New England Journal of Medicine*, 357, 370–79.
- Duan, Naihua, Willard G. Manning, Carl N. Morris, and Joseph P. Newhouse (1983), "A Comparison of Alternative Models for the Demand for Medical Care," *Journal of Business and Economic Statistics*, 1(2), 115–26.
- Fairfield, Joshua A.T. (2005), "Virtual Property," *Boston University Law Review*, 85, 1048–1102.
- Farrow, Harmonie and Y. Connie Yuan (2011), "Building Stronger Ties With Alumni through Facebook to Increase Volunteerism and Charitable Giving", *Journal of Computer-Mediated Communication*, 16(3), 445–64.
- Ferrar, Emilio, Pasquale De Meo, Giacomo Fiumara, and Alessandro Provetti (2012), "The Role of Strong and Weak Ties in Facebook: A Community Structure Perspective," *Proceedings of Computational Approaches to Social Modeling (ChASM)*.
- Gelman, Andrew and Jennifer Hill (2007), *Data Analysis Using Regression and Multilevel / Hierarchical Models*, Cambridge University Press.
- Granovetter, Mark S. (1973), "The strength of weak ties," *American Journal of Sociology*, 78(6), 1360-1380.
- Greengard, Samuel (2011), "Social Games, Virtual Goods," *Communication of the ACM*, 54(April), 19–21.
- Guadagni, Peter M. and John D. C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2(3), 203–38.



- Hartmann, Wesley R., Puneet Manchanda, Harikesh Nair, Matthew Bothner, Peter Dodds, David Godes, Kartik Hosanagar, and Catherine Tucker (2008), "Modeling Social Interactions: Identification, Empirical Methods and Policy Implications," *Marketing Letters*, 19, 287–304.
- Hoff, Peter D. (2003), "Random Effects Models for Network Data," Working Paper.
- Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock (2002), "Latent Space Approaches to Social Network Analysis," *Journal of the American Statistical Association*, 97(460), 1090–98.
- Iacobucci, Dawn (1990), "Derivation of Subgroups from Dyadic Interactions," *Quantitative Methods in Psychology*, 107(1), 114–32.
- Iacobucci, Dawn and Nigel Hopkins (1992), "Modeling Dyadic Interactions and Network in Marketing," *Journal of Marketing Research*, 29(February), 5–17.
- Leenders, Roger (2002), "Modeling Social Influence through Network Autocorrelation: Constructing the Weight Matrix," *Social Networks*, 24(January), 21–47.
- Lehonvirta, Vili (2009), "Virtual Item Sales as a Revenue Model: Identifying Attributes that Drive Purchase Decisions," *Electronic Commerce Research*, 9, 97–113.
- Ma, Liye., Ramayya Krishnan, and Alan Montgomery (2011), "Homophily or Influence? An Empirical Analysis of Purchase within a Social Network," Working Paper.
- Malone, Thomas W. (1981), "Toward a Theory of Intrinsically Motivating Instruction," *Cognitive Science*, 4, 333–69.
- Marsden, Peter V. and Karen E. Campbell (1984), "Measuring tie strength," *Social Forces*, 63(2), 482–501.
- McCulloch, Charles E. and Shayle R. Searle (2001), *Generalized, Linear, and Mixed Models*, John Wiley & Sons.
- McGonigal, Jane (2011), "Be a Gamer, Save the World," *Wall Street Journal*, January 21, 2011.
- McPherson, Miller, Lynn Smith-Lovin, and James M. Cook (2001), "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, 27:415–44.
- Ministry of Culture, People's Republic of China (2011), "The Annual Report on Chinese Online Video Game Industry, 2010".
- Moffitt, Robert A. (2001), "Policy Interventions, Low-Level Equilibria and Social Interactions," in *Social Dynamics*, S. Durlauf and H.P. Young, eds. Washington, DC: Brookings Institution Press/MIT Press, 45–82.
- Nair, Harikesh S., Puneet Manchanda, and Tulikaa Bhatia (2010), "Asymmetric Social Interactions in Physician Prescription Behavior: The Role of Opinion Leaders," *Journal of Marketing Research*, 47(October), 883–95.

Narayan, Vishal, Vithala R. Rao, and Carolyn Saunders (2011), "How Peer Influence Affects Attribute Preferences: A Bayesian Updating Mechanism," *Marketing Science*, 30(2), 368–84.

Ovide, Shira (2011), "Zynga's Revenue Growth Rate Is Slowing," *Deal Journal*, WSJ Blogs, online source: <http://blogs.wsj.com/deals/2011/09/21/zyngas-revenue-growth-rate-is-slowing/>

Rabe-Hesketh, Sophia and Anders Skrondal (2010), "Generalized Linear Mixed Models," *Elsevier Second Proof*, Elsevier.

Reeves, Byron and J. Leighton Read (2009), "The Game Tsunami Who Plays, How Much, and Why," *Total Engagement: Using Games and Virtual Worlds to Change the Way People Work and Businesses Compete*, Harvard Business Press.

SAS (2011), *SAS/STAT® 9.3 User's Guide*, SAS Institute Inc.

Schmidt, Dale (2012), "Under Control: Demand for Video Games Will Increase as Disposable Income Grows", IBISWorld Industry Report NN003 Video Games in the US, IBISWorld Inc.

Shriver, Scott K., Harikesh Nair, and Reto Hofstetter (2011), "Social Ties and User-Generated Content: Evidence from an Online Social Network," Working Paper.

Trusov, Michael, Anand V. Bodapati, and Randolph E. Bucklin (2010), "Determining Influential Users in Internet Social Networks," *Journal of Marketing Research*, 47(August), 643–58.

Wasserman, Stanley and Katherine Faust (1994), *Social Network Analysis: Methods and Applications*, New York: Cambridge University Press.

Williams, Dimitri, Nick Yee, Scott E. Caplan (2008), "Who Plays, How Much, and Why? Debunking the Stereotypical Gamer Profile", *Journal of Computer-Mediated Communication*, 13(4), 993–1018.

Wingfield, Nick (2011, September 9), "Virtual Products, Real Profits", *Wall Street Journal*, Eastern Edition. 9/9/2011, Vol. 258 Issue 59, pA1-A16. 2p. 1.

Xiang, Rongjing, Jennifer Neville, and Monica Rogati (2010), "Modeling Relationship Strength in Online Social Networks", *Proceeding WWW'10 Proceedings of the 19th international conference on World Wide Web*, Pages 981–90.

## CHAPTER 2

### MODELING CUSTOMERS' DEFECTION IN A SOCIAL NETWORK

#### *Abstract*

In this research we attempt to model customers' defection decisions within a social network. We propose a framework to jointly estimate a dyadic level tie strength model and an individual level defection decision model. We use tie strength model to reveal the latent, time-varying tie strength within each customer dyad. We then use the revealed tie strength states and customers' defection history to generate a set of social contagion covariates. Using these covariates we are able to capture the impact of strong ties and weak ties on customers' defection decisions. We test our framework with the data collected from an online game. We find that customers who actively interact with others tend to have strong ties with them. Also, customers with strong ties tend to have stronger influence on other customers' defections than customers with weak ties. Based on our model we propose an incremental impact measure that can be used to identify influential customers. We also propose a new approach to measure customers' social network value. To conclude this paper, we demonstrate the importance of correctly identifying and retaining influential customers with two policy simulations.

**Key Words:** customer relationship management, defection, social network

## 2.1 Introduction

It was in the early 1990s when companies started accepting that customers are their most valuable assets. As a result customer retention became a hot topic (Billington 1996). Researchers and practitioners painstakingly investigated the root causes of customer defection (Reichheld 1996). Interestingly, it was soon discovered that customers still defect even when they are satisfied, unless they are “completely satisfied” (Jones and Sasser 1995). Therefore researchers pay a lot of attention to the accurate measurement of customer satisfaction and model the relationship between satisfaction and defection (e.g., Bolton 1998, Chandrashekar et al. 2007).

However, analyzing customers’ defection behaviors is not an easy task. First, it is difficult to detect customers’ defections, especially in a non-contractual situation. Various methods have been developed to predict customers’ status (Fader and Hardie 2009). Most of these methods provide the probability distribution of whether customers are still “alive” or already “dead”. Yet with these outcomes it is still difficult to investigate how customers make their decisions to churn.

Second, social influence might play a role in customers’ defection, which is hard for researchers to capture. Researchers have found that satisfied and unsatisfied customers are likely to engage in word of mouth (Anderson 1998), and customers’ satisfaction levels are substantially related with the duration of customer-company relationships (Bolton 1998). But so far no one has put these two pieces together to show the impact of social influence on customer’s satisfaction and retention. An attempt has been made by Nitzan and Libai (2011). The authors use hazard model to analyze the impact of social network on customers churn from a telecommunication service provider. One important reason for this sparsity in literature lies in the lack of social connection and interaction data. Also, economists have already pointed that,

unless with detailed information of customers and observe customers' behaviors in a strictly controlled environment, it is difficult to capture the true social influence (Manski 1993).

Fortunately, data collected from online gaming offer researchers a unique opportunity to analyze customers' defection decisions within a social network. In online gaming data, researchers can observe how customers connect and interact with each other. Thus researchers can infer how customers influence each other through their social ties. More importantly, some games are running on a number of servers, each for only a limited period of time. Therefore, in these games a customer's whole life cycle is observable and her defection behaviors are also visible. Inspired by such a unique feature, in this research we model customers' defection decisions, using data collected from a game server.

The ability to model customer defection behaviors within a social network is of critical importance to companies. Knowing how customers influence each other's defection decisions, companies can identify influential customers and make more efforts to retain them. Further more, companies can leverage the social influence from these influential customers to retain more other customers. With models, companies are able to quantitatively measure the impact of one customer's defection decision on other customers' defection decisions. This incremental impact can be directly used to identify influential customer. This incremental impact, together with the purchasing records of those customers who are under influence, can be used to gauge each customer's "social influence value". A customer's social influence value, plus her intrinsic value (her own purchasing) can then be used to optimize a company's resource allocation for customer retention.

To model customers' defection behaviors within a social network, we propose a framework with a tie strength model and a defection decision model. The tie strength model uses

customers' interaction history to model each customer dyad's tie strength as a time-varying, binary, and latent state (strong vs. weak). With revealed tie strength states and customers' defection records, we generate four social contagion covariates. These covariates are used in the defection decision model to capture the "group-to-one" social influence on each individual customer's defection decision.

We jointly estimate the tie strength model and defection decision model using a MCMC algorithm. The results show that the more time two customers spend together, the more likely they have a strong relationship. The more recent their interactions are, the more likely they are strongly related. The results also show that customers with stronger ties tend to have stronger influence on each other's defection decision than those with weak ties.

With estimated parameters we measure the incremental impact of each customer's defection decision on other customers' decision to stay or leave. This measure of the incremental impact can be used to identify influential customers. The social network value, calculated with this incremental impact and customers' monetary contribution records can help a company to decide how to allocate their rendition efforts to customers. As a demonstration of the effectiveness of our framework, we use two policy simulations to show how a company can greatly increase its revenues by correctly identifying and retaining its influential customers.

To the best of our knowledge, this research is the first attempt to jointly model customers' tie strength and their defection decision within a social network. Different from the hazard model approach (e.g., Nitzan and Libai 2011), our defection decision model can help predict each individual customer's defection at each point of time. Our tie strength model can help identify influential customers from their interaction records. Our method can help companies improve their resource allocation to retain customers. Our findings from online gaming can be applied to

other settings as well, including physical social networks. This is because recent investigations show that customers have similar motivations and behaviors in the virtual world and the physical world (Bartle 2003; Malone 1981; Williams, Yee, and Caplan 2008).

The remainder of this paper is organized as follows: section 2 reviews the literature relevant to this research; section 3 describes the data used in this research. In section 4 we specify the models for tie strength and defection decision. In section 5 we present the estimation results from our models. In section 6 we discuss robustness tests, cross-validation, and policy simulation. We then conclude this paper with a discussion of the contributions and limitations of this research, and future research directions.

## **2.2 Literature Review**

Our research is relevant to three streams of literature: customer defection/retention, social influence, and social network value.

### *Customer Defection/Retention*

The topic of customer retention gained popularity in the early 90s, when companies became more customer-oriented (Billington 1996). The discussions focused on: (1) how to identify high value customers; and (2) how to allocate resources to retain valuable customers (Reinartz and Venkatesan 2008). Some researchers work on the relationship between customers' satisfaction and their loyalty to the firms (e.g., Bolton 1998, Shankar et al. 2003, Chandrashekar et al. 2007). Others directly investigate customers' defection or the duration of their relationship with firms.

Modeling customers' defection is relatively easier in the contractual setting, in which customers' defection can be detected by their termination of contracts or subscription. Neslin et al. (2006) run an interesting meta-analysis of the models used to predict customer churn. The

authors find that logit and probit model outperform other models.

Things get much more difficult in the non-contractual setting, when customers' defection cannot be directly observed. Hazard models are used to analyze the duration of customer-company relationship or customers' inter-purchase timing (Allenby et al. 1999, Seetharaman and Chintagunta 2003).

A variety of probability models have been developed to predict the probability of a customer's staying "alive" at a certain point of time. The two most widely used models are Pareto/NBD model (Schmittlein et al. 1987, Schmittlein and Peterson 1994) and Beta geometric/NBD model (Fader et al. 2005). Probability model are employed to predict customers' inter-purchase timing (e.g., Gupta 1991). Based on the output of these models, customer's life time value (CLV) can be estimated and companies' CRM decisions can be optimized (e.g., Reinartz and Kumar 2000, 2003).

Most literature in this stream treats each individual customer independently, ignoring the possible influence among customers. Recently, social network effects have already caught some CRM researchers' attention. For instance, Villanueva et al. (2008) find that customers acquired via word-of-mouth have much higher long-term value to the firm than the customers acquired through marketing activities. So far, literature is still sparse in the area of modeling customers' defection/retention with explicit consideration of social ties and social influence. The most recent attempt is carried out by Nitzan and Libai (2011). In this paper the authors use hazard model to investigate the impact of social network characteristics, tie strength, satisfaction, and economic incentives on customers' hazard to churn from a telecommunication service provider.



## *Social Influence*

In this research we try to model how customers influence each other's defection decisions through their social ties in an online social network. In particular, we want to investigate whether the impact on defection due to strong ties and weak ties is significantly different. In an online social network, customers form social connections by registration (e.g., becoming friends on the Facebook) or through interaction (e.g., sharing information, participating in activities, or sharing digital products). Interactions vary in scale and frequency. Some interactions are casual, which result in weak ties; while others are intense, which leads to strong ties. Accordingly, strong influence might be conveyed via strong ties. To establish this, we analyze a complete chain of "social interaction  $\rightarrow$  tie strength  $\rightarrow$  social influence" in our research.

So far most of the social influence literature has concentrated on the link of "social connection (interaction)  $\rightarrow$  social influence", particularly: (1) the spread of certain phenomena or behaviors within a social network (social contagion); (2) the influence from influential customers and opinion leaders. Examples of the former topic include the spread of obesity (Christakis and Fowler 2007) and happiness (Fowler and Christakis 2008); the diffusion of new products (Bhatt et al. 2010, Iyengar et al. 2011, Katona et al. 2011, Banerjee et al. 2012); the structure effect on the spread of behaviors (Centola 2010). The examples of the latter topic include: the asymmetric influence between opinion leaders and followers (Nair et al. 2010) and the identification of influential customers from their social behaviors (Trusov et al. 2010, Aral and Walker 2012).

There are a few papers discussing the link of "social interaction  $\rightarrow$  tie strength". Iacobucci and Hopkins (1992) model tie strength at discrete levels. Xiang et al. (2011) model tie strength as a continuous, latent covariate. We believe that tie strength, as a function of customers' social connections and interactions (usually observable), can help us identify influential

customers (usually unobservable). In Chapter 1, we “assign” tie strength levels (weak/strong) to each customer dyad at each time point, based on their social connections and interaction. We then generate social effect covariates using tie strength information and analyze strong ties and weak ties’ social influence on customers’ purchasing behaviors.

In this paper we adopt a similar approach. We first treat customers’ tie strength as latent and model it as a function of their interaction history, then use tie strength information to generate a set of social contagion covariates. Using these social contagion covariates we can investigate strong ties and weak ties’ social influence on customers’ defection decision. In other words, we utilize the complete chain of “social connection/interaction → tie strength → social influence” in this research.

#### *Social Network Value*

The social network value of a customer can be defined as “the expected profit from sales to other customers she may influence to buy, the other customers those may influence, and so on recursively” (Domingos and Richardson 2001, page 1). The literature in this stream is surprisingly sparse. One possible reason for this scarcity is the lack of data: it is difficult to piece together all the information needed to calculate the social network value (e.g., social connection, social interaction, purchasing... etc.) Another (and maybe a more important) reason is that the contexts of research varies dramatically from one social network to another (Facebook, Twitter, e-commerce, online gaming ... etc.). Apparently it is difficult to build a uniform framework to analyze customers’ social network value across so diversified contexts.

One attempt is made by Gupta et al. (2009). In their paper the authors use a structural model to estimate the value of direct network effects (“buyer-buyer”) and indirect network effects (“buyer-seller”). However, the value of networked effects estimated in this paper is still

not the value of social influence among customers.

Data mining researchers also make some progress in this area. Modeling customer social network as a Markov random field, Domingos and Richardson (2001) first assume that all customers are in isolation and estimate the value contribution of each customer (“intrinsic value”). They then assume that there exists social influence among customers and estimate the value contribution (“total value”). They define the difference between a customer’s total value and intrinsic value as her “social network value”. This approach is suitable for the scenarios of new customer acquisition or the adoption of new product. It is still difficult to capture customers’ social network value in a repeated purchasing context.

In this research we propose a different approach to measure customers’ social network values. We first model customers’ defection decisions. Then we estimate the incremental impact of each customer’s defection decision on other customers’ decision to stay. Combining this incremental impact and the purchasing records of those influenced customers, we estimate the expected incremental value contribution of those customers. This approach can be used as a measure of the value of a customer’s social influence on other customers.

## **2.3 Data**

The data used in this research were collected from an online video game server in Shanghai, China. Online games, especially massively multiplayer online role-playing games (MMORPGs), are often run on different servers. On each server the game runs for several months (six months in our case). After that the game will start all over again. By doing so, company can attract a continuous inflow of new customers; because when those newcomers join the new round of the game they won’t be lagged too far behind.

Such a way of game operation affords us a unique opportunity to observe customers' defection behaviors within a social network. Even though customers haven't signed contracts or subscribed to the game, we can still observe their exit from the game. When customers register into a game, they can choose to join different servers (at different stages of the game). They will be informed about the closing date of the server. And as the closing date approaches, they will see the message every time they log on the game to remind them to make plans for the ending of the game. In this situation, if we observe that a customer never came back till the end of the game, we can safely claim that this customer left for good after her final logon. Therefore, simple but well-performed logit and probit models can be employed to analyze customers' defection decisions (Neslin et al. 2006).

Comparing with other social network data, online gaming data have several additional attractive features. For instance, in online gaming data customers' social ties can be observed. Their social interactions are painstakingly recorded. Their purchasing behaviors are also well logged. All these advantages make it possible for researchers to infer the social influence among customers. Also, recently large surveys on game players show that customers have very similar motives and behaviors in the virtual world and the physical world (Williams et al. 2008, Lehonvirta 2009). This means what we find in an online social network can be applied to an offline social network as well.

The game recorded in our data ran for a period of six months. Unfortunately, the database only recorded the most recent four month worth of behavioral data on a rolling basis.<sup>27</sup> Therefore we can only observe customers' interactions and purchases during the final four months. The first two months' data are missing. In other words, we have a left truncation issue. All customers'

---

<sup>27</sup> The system records all gamers' activities (accurate down to a second), therefore the data recorded everyday could be overwhelming. To save storage space, the system only records the most recent four months' data.

registration information, such as virtual demographic and friendship information, is kept in a separate database. Customers' login information is also complete, therefore we can observe each customer's complete life cycle, including when they left the game for good.

Most customers (more than 75%) in the game only showed up once or twice, and then disappeared. Therefore, we resample the data and keep the records of regular customer only. Based on our discussions with company managers, a "regular" customer is defined as the customer who: (1) stayed in the game for at least one month; (2) interacted with at least one other customer for at least one month in the game. We have 181 customers in our final sample. These customers form 1,315 friendship dyads (dyads that registered as friends) and 1,262 interacting dyads (dyads that interacted at least once).

In game customers can take a variety of actions. First, customers can become friends with each other, in a Facebook manner. Second, customers can participate in various missions, independently, or jointly with other customers. Third, customers can buy (sell) virtual products from (to) virtual merchants or other customers in the game. Some virtual products are priced with virtual currency, while others are priced with real money. Forth, customers can share virtual products with each other, free of charge. Fifth, customers can build or join different online groups (in gaming jargon, these groups are called "guilds"). Table 2.1 presents some descriptive statistics of the customers:

In this research, we can identify two types of social connections: friendship and partnership. If two customers register as friends, they have friendship connection between them. If two customers interact with each other (e.g., participating missions together, trading or sharing virtual products with each other ... etc.), they are involved in a partnership and they become partners. Friendship starts with registration and ends with one friend's leaving. Partnership starts

with first interaction and ends with one partner's defection. These two social relationships usually do not coexist in the same dyads. As a matter of fact, in our data they only have a small overlap (Table 2.2).

**Table 2.1**  
**Descriptive Statistics of Sampled Regular Customers**

Variable	min	max	mean	median	std dev
Number of Friends	0.00	123.00	25.01	19.00	23.60
Number Guilds Joined <sup>28</sup>	0.00	11.00	2.72	2.00	2.16
Number of Interacting Dyads (In-Degree)	0.00	27.00	2.22	1.00	2.79
Number of Interacting Dyads (Out-Degree)	0.00	21.00	2.22	1.00	2.67
Online Time (Hours per Week)	0.00	167.55	31.89	13.09	41.71
Number of Missions Accomplished per Week	0.00	533.00	19.17	1.00	47.35
Tenure (Weeks in Game Since Registration)	3.00	24.00	13.62	13.00	6.34
Dyad Duration (Week)	1.00	18.00	3.24	1.00	3.51
Weekly Interacting Frequency (Interacting Customers)	1	472	4.64	2	11.70
Weekly Cash Payment (Paying Customers, RMB) <sup>29</sup>	2	13,064	745.50	294	1,567.33

**Table 2.2**  
**Overlap between the Interacting Network and the Friendship Network**

Friendship	Partnership (Interacting Relationship)		
	No	Yes	Sub Total
No	0 0.00%	669 33.72%	669 33.72%
Yes	722 36.39%	593 29.89%	1,315 66.28%
Sub Total	722 36.39%	1,262 63.61%	1,984 100.00%

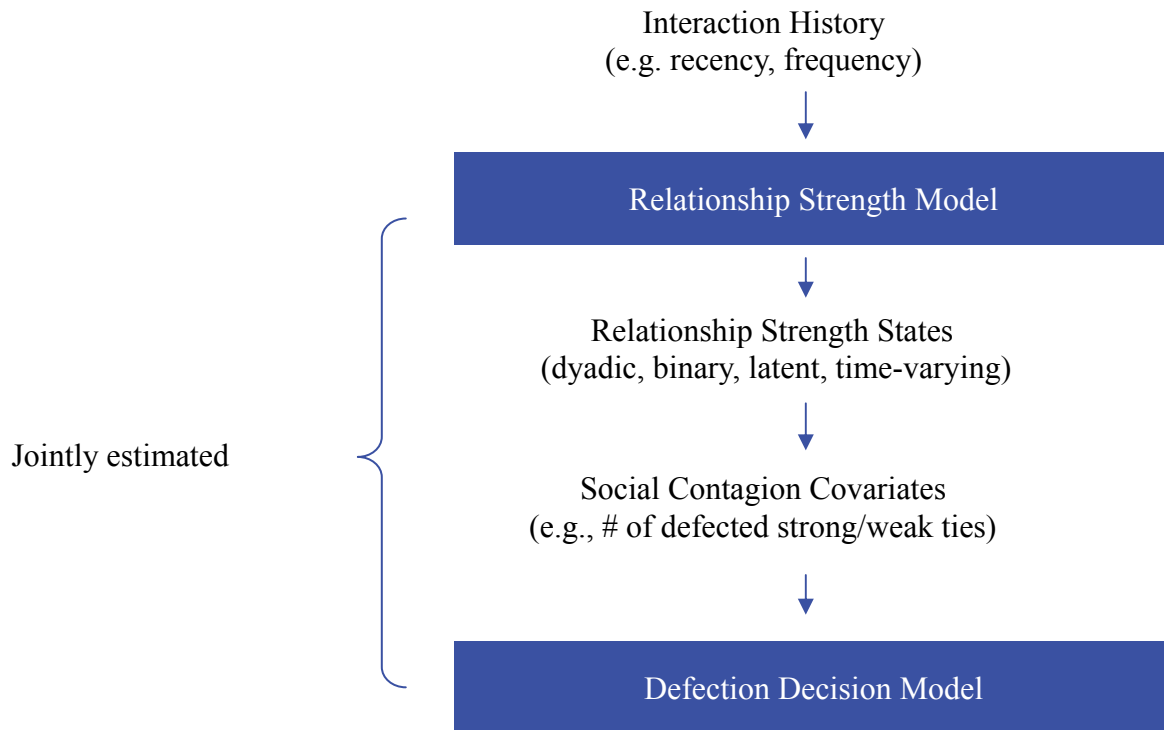
<sup>28</sup> We deleted all of the guilds with less than five members.

<sup>29</sup> 1RMB  $\approx$  0.16USD

## 2.4 Models

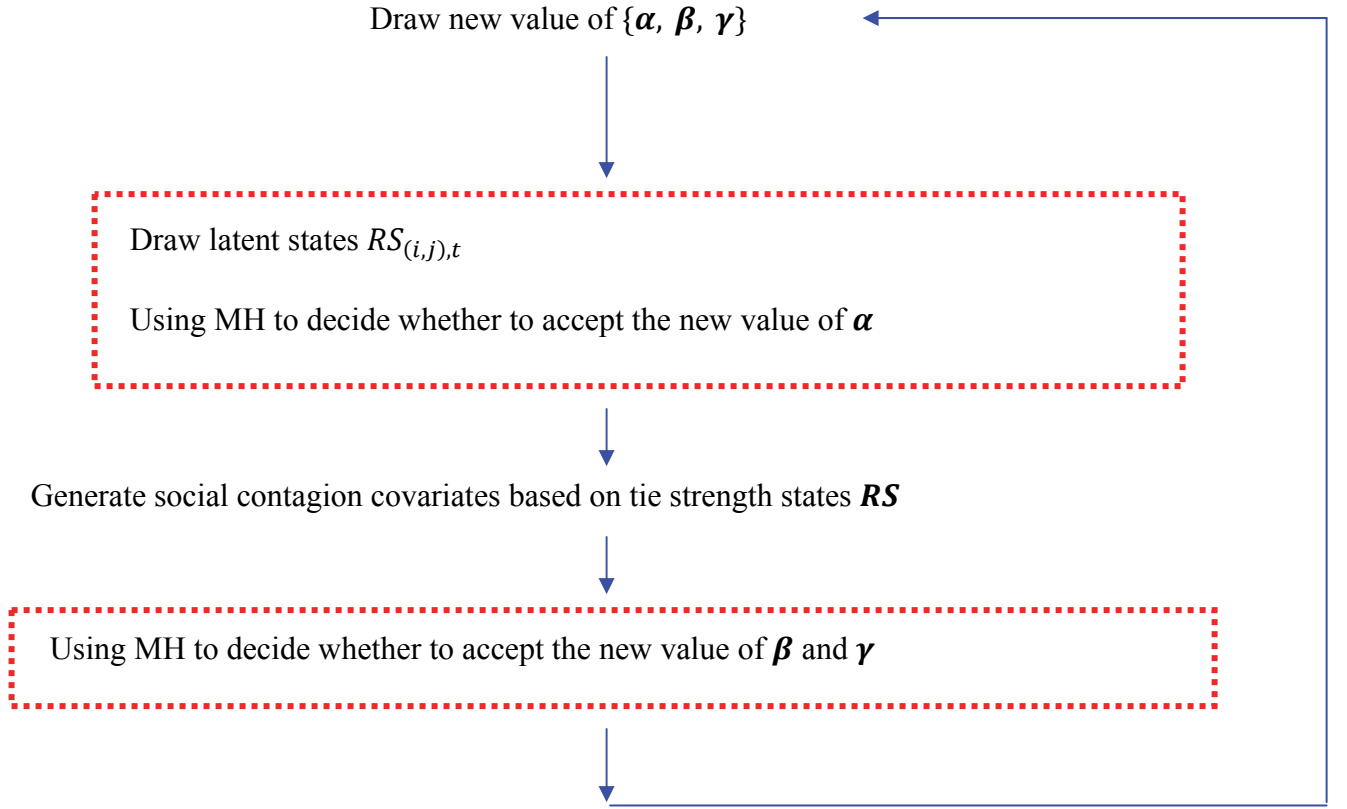
In this research we build a framework with two models: a tie strength model, and a defection decision model (Figure 2.1).

**Figure 2.1**  
**The Research Framework**



In tie strength model we use each customer dyad's interaction history to reveal their tie strength (time-varying and latent). Combining tie strength information and customers' defection history we are able to generate a set of social contagion covariates. Then in defection decision model we use these social contagion covariates to capture strong ties and weak ties' social influence on a customer's defection decision. A Markov Monte Carlo Chain (MCMC) algorithm is used to iterate between these two models until the chain converges. See Figure 2.2.

**Figure 2.2**  
**The MCMC Algorithm**



After these two models are estimated we then use the parameters to measure each customer's incremental impact on other customers' defection decisions. This incremental impact measure can be directly used to identify influential customers. Combining this incremental impact measure with customers' purchasing record we are able to gauge each customer's social network value.

#### *Tie Strength Model*

We use each customer dyad's interaction history to model their tie strength as a two-state (strong/weak), time-varying latent state. We decide to use a specific type of interactions in this dyadic level model: customers' sharing behaviors, because it is a good indicator of the nature of the relationship between two customers.

As mentioned in the Data section, virtual products can be bought, sold and shared among



customers. We find that most (more than 90%) of such transactions among customers are free of charge. Unlike other interactions such as mission participation, sharing behavior is the *direct* interaction between the giver and receiver.<sup>30</sup> It is also well recorded. More importantly, the items shared among customers have values (bought with game currency or real money). Therefore, sharing among two customers can be a good sign of an affectionate relationship.

In this model we treat customers' relationship as *nondirectional*, because in our prior project we find that in 89% cases customer dyads have a symmetric relationship (see Chapter 1). Therefore, the interaction between customer  $i$  and  $j$  in week  $t$  is calculated as:  $\frac{n_{i \rightarrow j,t} + n_{j \rightarrow i,t}}{2}$ , where  $n_{i \rightarrow j,t}$  is the frequency customer  $i$  giving to customer  $j$  in week  $t$ .

We use a probit model to describe each customer dyad's tie strength as a binary, time-varying latent state:

$$P(s_{(i,j),t} = 1 | \mathbf{X}_{(i,j),t}^{Hist}) = G(\mathbf{X}_{(i,j),t}^{Hist} \cdot \boldsymbol{\alpha}), \quad (2.1)$$

where  $s_{(i,j),t}$  is customer  $i$  and  $j$ 's tie strength in week  $t$  (1: strong tie; 0: weak tie);  $G(\cdot)$  is the CDF of a normal distribution with its mean at  $\mathbf{X}_{(i,j),t}^{Hist} \cdot \boldsymbol{\alpha}$ . To make the model estimable, we fix the variance of this normal distribution to 1;  $\mathbf{X}_{(i,j),t}^{Hist}$  is the vector of customer  $i$  and  $j$ 's interaction history up till week  $t$ ;  $\boldsymbol{\alpha}$  is the vector of the corresponding parameters. We use following parameters in this model:

1) *Partner Similarity*. This is the measure of the overlap between customer  $i$  and  $j$ 's partners

(customers with interacting relationship). It is calculated as  $\frac{NP_{(i,j),t}}{NP_{it} + NP_{jt}}$ , where  $NP_{(i,j),t}$  is the

number of the "mutual partners" between customer  $i$  and  $j$  in week  $t$ ;  $NP_{it}$  and  $NP_{jt}$  are the

---

<sup>30</sup> In our data we can observe that the two customers show up in the same mission at the same time, but we do not have direct evidence that these two customers are interacting with each other. As matter of fact, they might not see each other in the game at all.

number partners of customer  $i$  and  $j$  have in week  $t$ . This measure varies from 0 (no overlap) to 0.5 (full overlap);

- 2) *Online Time Overlap*. This is the measure how often customer  $i$  and  $j$  log in the game simultaneously. It is defined as  $\frac{OT_{(i,j),t}}{OT_{it}+OT_{jt}}$ , where  $OT_{(i,j),t}$  is the length of time when customer  $i$  and  $j$  are both in the game in week  $t$ ;  $OT_{it}$  and  $OT_{jt}$  are the time customers  $i$  and  $j$  spend in week  $t$ . Because in the data we cannot observe customers direct interactions in the missions, we use this measure as proxy of the interaction intensity between customer  $i$  and  $j$ . This measure varies from 0 (no overlap) to 0.5 (full overlap);
- 3) *Cumulative Count of Interactions*. This covariate counts the total number of interactions since the beginning of customer  $i$  and  $j$ 's partnership. It is used to track customer  $i$  and  $j$ 's interaction history;
- 4) *Recency of Interaction*. This covariate records the time lapse since customer  $i$  and  $j$ 's last interaction. It is used to capture the time diminishing impact on customers' tie strength.

#### *Social Contagion Covariates*

Basically, customers' defection process can be treated as the "spread of defecting behaviors" on a social network. This spreading process is similar to the adoption process of a new product (in a sense, defection can be considered as a "reverse-adoption"). So far, a large number of social influence papers have demonstrated social contagion in new product diffusion over a social network (e.g., Nair et al. 2010, Iyengar et al. 1011). In this research we use a similar approach to capture the social contagion effect in customers' defection behaviors.

However, the mechanism of social contagion is more complicated in a defection scenario than it is in a new product adoption scenario. This is because when a customer considers her defection decision, she is under influence from two groups: (1) her partners who have already

left; (2) her partners who have chosen to stay. The former will increase her probability to leave, while the latter will reduce that probability. To capture social contagion in this more complex setting, we need to employ a set of social contagion covariates.

One simple approach is to count the number of left partners and staying partners, using them as the social contagion covariates in the defection decision model. With this approach we can capture social contagion using a single model. The model can be estimated with MLE. In fact, we use this specification as one of our benchmark models. This approach, however, cannot help us differentiate customers by their influence, not to mention identify influential customers.

To be able to identify influential customers with their *observable* interaction behaviors, in our framework we investigate the complete link of “social interaction  $\rightarrow$  tie strength  $\rightarrow$  social influence”. In order to do so, we use the latent tie strength information and customers’ defection records to construct a set of four social contagion covariates:

- 1)  $SS_{it}$ : the number of customer  $i$ ’s staying partners in week  $t$ , with strong ties;
- 2)  $SW_{it}$ : the number of customer  $i$ ’s staying partners in week  $t$ , with weak ties;
- 3)  $CDS_{it}$ : the *cumulative* number of customer  $i$ ’s defected partners, with strong ties (in the week when they left);
- 4)  $CDW_{it}$ : the *cumulative* number of customer  $i$ ’s defected partners, with weak ties (in the week when they left).

These covariates are used in the defection decision model. To construct these four covariates, in each week we actually categorize each individual customer’ partners (current and left) into  $2 \times 2 = 4$  groups (“strong tie” vs. “weak tie”, and “staying” vs. “left”). By definition, unless these covariates have very small variation (which is not the case in this research) multicollinearity will not be an issue in our estimation. As a test, we calculate the correlation among these four

covariates at a number of iterations. We do not observe significant correlations among the covariates.

**Table 2.3**  
**Scenarios of Incremental Effects**

Scenario	Customer $i$ 's Decision	Customer $i$ 's Tie Strength with $j$	Social Contagion		Incremental Effect on Customer $j$ 's Probability to Stay
			Sign	Scale	
1	Stay	Strong	Negative	Large	Large and positive effect
2	Stay	Weak	Negative	Small	Small and positive effect
3	Leave	Strong	Positive	Large	Large and negative effect
4	Leave	Weak	Positive	Small	Small and negative effect

#### *Defection Decision Model*

We analyze each individual customer's defection decision in each week with an individual level, logit model:

$$P(d_{it} = 1 | \mathbf{s}_{it}, \mathbf{X}_{it}^{SE}, \mathbf{X}_{it}^{SC}) = F\left( \underbrace{\mathbf{X}_{it}^{SE'} \cdot \boldsymbol{\beta}}_{\text{self effects}} + \underbrace{\mathbf{X}_{it}^{SC'} \cdot \boldsymbol{\gamma}}_{\text{social contagion}} \right), \quad (2.2)$$

where  $d_{it}$  is customer  $i$ 's defection decision in week  $t$  (1: leave; 0: stay);  $\mathbf{X}_{it}^{SE}$  is the vector of self effects covariates;  $\mathbf{X}_{it}^{SC} = [SS_{it}, SW_{it}, CDS_{it}, CDW_{it}]$  is the vector of four aforementioned social contagion covariates;  $\boldsymbol{\beta}$  and  $\boldsymbol{\gamma}$  are the vectors of coefficients.  $\mathbf{X}_{it}^{SE}$  includes following covariates:

- 1) *Count of Guilds Attended*. This covariate is the number of online groups customer  $i$  has joined up till week  $t$ . It is used to measure a customer's willingness to participate in group activities;
- 2) *Tenure in the Game*. This covariate is the number of weeks customer  $i$  has already stayed in the game since registration. It is used to track the trend of defection with time;

- 3) *ln(Online Time)*. This is the logarithm of customer's online time (in hours) in week  $t$ . It is a measure of customer  $i$ 's activeness;
- 4) *ln(Cumulative Online Time)*. This is the logarithm of customer's cumulative online time (in hours) up till week  $t$ . It is used to track the trend of customer defection with time;
- 5) *Count of Missions*: This the number of missions customer  $i$  joined in week  $t$ . It is also a measure of customer  $i$ 's activeness;
- 6) *Change in Level*: This is customer  $i$ 's change in performance level in week  $t$  ( $Level_t - Level_{t-1}$ ). It is used to measure customer  $i$ 's achievement in week  $t$ ;
- 7) *Cumulative Payments*: This is the (real) money customer  $i$  has paid up till week  $t$ . We want to use this covariate to check whether customers experience "escalated commitments" (the more they have spent, the less willing they are to give up the game). Note that in this particular game, when a customer exists, she cannot get any refund. What she has spent will stay in the game. Therefore, we are expecting to observe some escalated commitment effect.

In our data, the time series of a customer's defection decision  $\{d_{it}\}$  is:  $\{0, \dots, 1\}$  (customer defected) or  $\{0, \dots, 0\}$  (customer never defected). Therefore the likelihood function is:

$$L = \prod_{i=1}^N \prod_{t=T_i^0}^{T_i^1} \{P(d_{it} = 1)^{d_{it}} \cdot P(d_{it} = 0)^{1-d_{it}}\}, \quad (2.3)$$

where  $N$  is the total number of customers;  $[T_i^0, T_i^1]$  is the observation period of customer  $i$ .

### *Incremental Impact and Social Network Value*

After we estimate the defection decision model, we are able to estimate the incremental impact of customer  $i$ 's defection decision (stay, leave) on her partner  $j$ 's decision to stay. There are four scenarios. Note that we are estimating the impact on customer  $j$ 's probability to *stay* (i.e.,  $P(d_{jt} = 0)$ ).

*Scenario 1.* Customer  $i$  decides to *stay* in week  $t$ , the incremental impact of her stay on

her *strongly*-tied partner  $j$  is:

$$\Delta_{\langle i,j \rangle,t} = \frac{\partial P(d_{jt} = 0 | \mathbf{X}_{jt}^{SC})}{\partial SS_{jt}} = -f(\mathbf{X}_{jt}^{SE} \cdot \boldsymbol{\beta} + \mathbf{X}_{jt}^{SC} \cdot \boldsymbol{\gamma}) \times \gamma_1, \quad (2.4)$$

where  $\mathbf{X}_{it}^{SC} = [SS_{it}, SW_{it}, CDS_{it}, CDW_{it}]$  is the vector of social contagion covariates.

*Scenario 2.* Customer  $i$  decides to *stay* in week  $t$ , the incremental impact of her stay on her *weakly*-tied partner  $j$  is:

$$\Delta_{\langle i,j \rangle,t} = \frac{\partial P(d_{jt} = 0 | \mathbf{X}_{jt}^{SC})}{\partial SW_{jt}} = -f(\mathbf{X}_{jt}^{SE} \cdot \boldsymbol{\beta} + \mathbf{X}_{jt}^{SC} \cdot \boldsymbol{\gamma}) \times \gamma_2. \quad (2.5)$$

*Scenario 3.* Customer  $i$  decides to *leave* in week  $t$ , the incremental impact of her defection on her *strongly*-tied partner  $j$  is:

$$\Delta_{\langle i,j \rangle,t} = \frac{\partial P(d_{jt} = 0 | \mathbf{X}_{jt}^{SC})}{\partial CDS_{jt}} = -f(\mathbf{X}_{jt}^{SE} \cdot \boldsymbol{\beta} + \mathbf{X}_{jt}^{SC} \cdot \boldsymbol{\gamma}) \times \gamma_3. \quad (2.6)$$

*Scenario 4.* Customer  $i$  decides to *leave* in week  $t$ , the incremental impact of her defection on her *weakly*-tied partner  $j$  is:

$$\Delta_{\langle i,j \rangle,t} = \frac{\partial P(d_{jt} = 0 | \mathbf{X}_{jt}^{SC})}{\partial CDW_{jt}} = -f(\mathbf{X}_{jt}^{SE} \cdot \boldsymbol{\beta} + \mathbf{X}_{jt}^{SC} \cdot \boldsymbol{\gamma}) \times \gamma_4. \quad (2.7)$$

Note that subscript  $\langle i, j \rangle$  indicates that the incremental impact is *directional* (from  $i$  to  $j$ ). Recall that  $SS_{jt}$ ,  $SW_{jt}$ ,  $CDS_{jt}$  and  $CDW_{jt}$  are the social contagion covariates of customer  $j$  in week  $t$ ,  $\gamma_1, \dots, \gamma_4$  are the coefficients of these four social contagion covariates.

Apparently,  $\Delta_{\langle i,j \rangle,t}$  can be used to locate influential customers. If we find a customer has a large number of social ties, most of them carry high incremental impacts, she can be labeled as an influential customer.

In data, we can also observe customer  $j$ 's monetary contribution in week  $t$ :  $\pi_{jt}$ . Thus the expected incremental lift on customer  $j$ 's monetary contribution, due to customer  $i$ 's defection

decision in week  $t$ , is:

$$SV_{\langle i,j \rangle, t} = \Delta_{\langle i,j \rangle, t} \times \pi_{jt}. \quad (2.8)$$

$SV_{\langle i,j \rangle, t}$  can therefore be used as the measure of the value of customer  $i$ 's incremental social influence on customer  $j$  in week  $t$ . The total value of customer  $i$ 's incremental social influence is:

$$SV_i = \begin{cases} \sum_{j \in SP(i)} \sum_{t=T_{(i,j)}^1}^{T_{(i,j)}^2} SV_{\langle i,j \rangle, t} & , \text{ before } i\text{'s defection} \\ \sum_{j \in SP(i)} \sum_{t=T_{(i,j)}^2+1}^{T_{(i,j)}^3} SV_{\langle i,j \rangle, t} & , \text{ after } i\text{'s defection} \end{cases} \quad (2.9)$$

where  $SP(i)$  is the set of customer  $i$ 's partners;  $T_{(i,j)}^1$  is the time when customer  $i$  and  $j$  start interaction;  $T_{(i,j)}^2$  is the time when customer  $i$  defects;  $T_{(i,j)}^3$  is the time when customer  $j$  defects (both customers leave the game).

$SV_i$  can be defined as a measure customer  $i$ 's social network value. Note that here we only measure the value of *direct* (first order) influence between two customers. In other words, we do not evaluate the influence on partners' partner. For example, if we observe a partnership of " $i \rightarrow j \rightarrow k$ " (there is no interaction between  $i$  and  $k$ ), we only measure the value of influence from  $i$  to  $j$ . We do not measure the value of influence from  $i$  to  $k$  through  $j$ . Therefore,  $SV_i$  is a *conservative* measure of customer  $i$ 's social network value.

## 2.5 Estimation and Results

### *Estimation Algorithm*

To jointly estimate the tie strength model and the defection decision model, we use an MCMC algorithm. Because the posterior distribution does not have a closed form, we employ the random walk Metropolis-Hastings algorithm to update the parameter estimates in iterations

(see Appendix).

For each parameter, we use a diffuse normal prior  $N(0,100)$ . To generate the initial value for parameters, we take the following steps:

- 1) *Tie Strength States*. Because tie strength states are latent, we use the level of each customer dyad's interaction frequency as a proxy. In each week, if a dyad's interaction frequency is higher (lower) than a cutoff value, we label this dyad's tie strength as "strong" ("weak"). We use three sets of cutoff values: 50 percentile, 70 percentile, and 80 percentile of weekly interactions. Thus we get three sets of tie strength states;
- 2) *Tie Strength Model Parameters*. After we get the initial values of tie strength states, we then estimate three probit models using these three sets of tie strength initial values. We use the three sets of MLE estimates from these models as the initial values of tie strength model parameters;
- 3) *Defection Decision Model Parameters*. We use assigned tie strength states and customers' defection records to generate the four social contagion covariates:  $SS_{it}$ ,  $SW_{it}$ ,  $CDS_{it}$ , and  $CDW_{it}$ . We then estimate the three individual level, logit models of customers' defection decisions. We use these three sets of MLE of parameters as the initial values for the defection decision model.

With these three different sets of initial values, we are able to run three Markov chains simultaneously. Because latent states are involved in the model, the chains converge slowly. We run the chains for 206,000 iterations and record every tenth draw. We tune the random-walk steps several times during iterations to control the acceptance rate to a level between 25% and 50% (Gelman et al. 2004). The average acceptance rate of these three chains is approximately 30%. We then use the final 60,000 draws to calculate the posterior statistics. We use the posterior



distribution of these three chains to check whether they converge to the same position. We follow Gelman et al. (2004)'s approach to calculate the “potential scale reduction”  $\hat{R}$ . All 17 parameters'  $\hat{R}$ s range from 1.00 to 1.03, which is all very close to 1 (below the 1.1 criterion suggested by Gelman et al. 2004). This provides us strong evidence that the chains do converge.

### *Estimation Results*

The estimation results are given in Table 2.4.

**Table 2.4**  
**Estimation Results of Tie Strength Model and Defection Decision Model**

Model	Parameter	Expected Sign	Covariate	2.5%	50.0%	97.5%	
Tie strength	$\alpha_0$	+/-	Intercept	-25.45	-9.36	-3.25	*
	$\alpha_1$	+	Partner Similarity	-17.11	-1.85	15.24	
	$\alpha_2$	+	Online Time Overlap	5.49	12.46	24.29	*
	$\alpha_3$	+/-	Cumulative Count of Interactions	-11.99	-5.14	2.14	
	$\alpha_4$	-	Recency of Interaction	-15.61	-6.16	-1.25	*
Defection Decision	$\beta_0$	+/-	Intercept	-6.20	-4.51	-2.84	*
	$\beta_1$	-	Count of Guilds Attended	-0.35	-0.21	-0.07	*
	$\beta_2$	+/-	Tenure in the Game	0.05	0.09	0.13	*
	$\beta_3$	-	$\ln(\text{Online Time})$	-0.30	-0.01	0.26	
	$\beta_4$	+/-	$\ln(\text{Cumulative Online Time})$	0.03	0.18	0.34	*
	$\beta_5$	-	Count of Missions	-0.17	-0.07	-0.02	*
	$\beta_6$	-	Change in Level	-0.08	-0.05	-0.02	*
	$\beta_7$	+/-	Cumulative Payments	0.00	0.05	0.10	*
	$\gamma_1$	-	Count of Stayed Strong Ties ( <i>SS</i> )	-2.52	-1.31	-0.65	*
	$\gamma_2$	-	Count of Stayed Weak Ties ( <i>SW</i> )	-0.09	-0.04	-0.01	*
	$\gamma_3$	+	Cum. Count of Defected Strong Ties ( <i>CDS</i> )	0.43	1.62	3.59	*
	$\gamma_4$	+	Cum. Count of Defected Weak Ties ( <i>CDW</i> )	0.01	0.06	0.11	*
<i>LL</i>			Log likelihood	-363.74	-356.43	-351.45	

As we can see, two parameters are significant in the tie strength model: *Online Time Overlap* and *Recency of Interaction*. Online time overlap has a positive sign, which means the more time two customers spend together, the more likely they have strong relationship. Recency of interaction (weeks since last interaction) has a negative sign, which means the more recent

two customers' last interaction is, the more likely they maintain a strong relationship.

Most of the parameters in the defection decision model are significant. Among self-effect parameters ( $\beta$ 's), we find that (note that this model predicts a customer's probability to *leave*):

- 1) *Count of Guilds Attended* is significantly positive. This tells us that if a customer is actively involved in various online groups, she is less likely to leave;
- 2) *Tenure in the Game* is positive. This means when a customer stays for a long time in the game, she has a high probability to defect;
- 3) *ln(Cumulative Online Time)* is positive. It means that we do not discover “escalated commitment” phenomenon in this case. The more time a customer spends on the game, the more likely she will leave;
- 4) *Count of Missions* is negative. This parameter is a measure of a customer's activity in the game. The more active a customer is, the less likely she will defect;
- 5) *Change in Level* is negative. This means if a customer is advancing in her performance level she is less likely to quit;
- 6) *Cumulative Payment* is positive. Again, we do not find “escalated commitment” phenomenon in this case. The more money a customer has spent on the game, the more likely she is leaving.

We get three interesting findings in four social contagion parameters ( $\gamma$ 's).

First, we find that *Count of Stayed Strong Ties* ( $SS_{it}$ ) and *Count of Stayed Weak Ties* ( $SW_{it}$ ) both have negative signs. This tells us that if a customer decides to stay, the customers who are interacting with her will become less likely to leave;

Second, we find that *Cumulative Count of Defected Strong Ties* ( $CDS_{it}$ ) and *Cumulative Count of Defected Weak Ties* ( $CDW_{it}$ ) both have positive signs. Apparently, one customer's

leave will have impact on other customers who are interacting with her, making them more likely to leave as well. This reveals the social contagion factor in customers' defection behaviors;

Third, we also find that strong tie covariates have larger scale than weak tie covariates ( $|SS_{it}| > |SW_{it}|$ ,  $|CDS_{it}| > |CDW_{it}|$ ). This means customers with strong ties have stronger influence than customers with weak ties.

### *Incremental Impact on Defection Decision*

The incremental impact of one customer's defection decision (stay or leave) on other customers' probability to stay can be calculated with equation (2.4) - (2.7). Our defection decision model is a logit model, therefore the first component of the incremental effect equation is the density of logistic distribution:

$$f(\mathbf{X}_{jt}^{SE} \cdot \boldsymbol{\beta} + \mathbf{X}_{jt}^{SC} \cdot \boldsymbol{\gamma}) = \frac{\exp(\mathbf{X}_{jt}^{SE} \cdot \boldsymbol{\beta} + \mathbf{X}_{jt}^{SC} \cdot \boldsymbol{\gamma})}{\{1 + \exp[(\mathbf{X}_{jt}^{SE} \cdot \boldsymbol{\beta} + \mathbf{X}_{jt}^{SC} \cdot \boldsymbol{\gamma})]\}^2}. \quad (2.10)$$

Density is always positive, therefore the sign and scale of incremental impact  $\Delta_{(i,j),t}$  depends on the sign and scale of social contagion covariates. Note that in the four scenarios listed in Table 2.3, customer  $i$ 's incremental impact on customer  $j$  depends on not only her own scenario ( $\gamma$ ), but also other customers' scenarios (embedded in  $\exp(\mathbf{X}_{jt}^{SE} \cdot \boldsymbol{\beta} + \mathbf{X}_{jt}^{SC} \cdot \boldsymbol{\gamma})$ ).

## **2.6 Tests, Cross-Validation, and Policy Simulation**

In this section we first test our models' goodness-of-fit against four benchmark models. Then we test our models' robustness to left truncation. We also cross-validate our models using a "Leave-One-Out" (LOO) approach. We finally run two policy simulations to demonstrate the importance of correctly identifying and effectively retaining influential customers.

### *Benchmark Models*

In our investigation of customers' defection decisions, if we do not want to explicitly model customers' tie strength, we have two options: (1) completely ignore the tie strength. We can simply count the number of partners staying or leaving in each week. We then use these counts as social contagion covariates; (2) use some rule of thumb to assign each customer dyad a tie strength level.

In each week we can assign each customer dyad into strong tie group or weak tie group, based on its weekly interaction frequency. We can choose some cutoff value, such as 50 percentile, 70 percentile, or 80 percentile. This is exactly our approach to generate initial value for MCMC algorithm. From option 1 and 2 we specify four benchmark models. The estimation results of these four models and our full model are given in Table 2.5.

In Table 2.5 Model *P00* is the one without considering tie strength; Model *P50*, *P70* and *P80* are the models using 50, 70 and 80 percentile as cutoff values to assign tie strength states. “*Joint*” model is our full model jointly estimated with the tie strength model. As we can see in this table, all five models have similar coefficient estimates, but *Joint* Model outperforms all four benchmark models in terms of log likelihood and Akaike Information Criterion (AIC).

### *Robustness Test*

Our data is left truncated; therefore we need to test our models' robustness to this issue. In addition to our original data set (including both truncated and non-truncated customers), we prepare three more datasets: (1) *T0*: all the dyads in this data set do not suffer from truncation issue; (2) *T1*: the dyads in this data set contain only one truncated customer; (3) *T2*: all dyads in this data set contain two truncated customers. Accordingly, we call our original, full data set *TA*. We then estimate our model on *T0* – *T2* and compare the results with the ones from *TA*.

**Table 2.5**  
**Estimate Results of Benchmark Models**

	Model									
	<i>P00</i>		<i>P50</i>		<i>P70</i>		<i>P80</i>		<i>Joint</i>	
Intercept	-4.17	(0.76)*	-4.16	(0.77)	-4.16	(0.76)	-4.16	(0.75)	-4.51	(-6.20,-2.84)**
Count of Guilds Attended	-0.16	(0.07)	-0.15	(0.07)	-0.14	(0.07)	-0.22	(0.06)	-0.21	(-0.35, 0.07)
Tenure in the Game	0.09	(0.02)	0.09	(0.02)	0.09	(0.02)	0.09	(0.02)	0.09	( 0.05, 0.13)
<i>ln</i> (Online Time)	-0.44	(0.10)	-0.40	(0.13)	-0.47	(0.12)	-0.46	(0.12)	-0.01	(-0.30, 0.26)
<i>ln</i> (Cumulative Online Time)	0.15	(0.07)	0.15	(0.07)	0.15	(0.07)	0.14	(0.07)	0.18	( 0.03, 0.34)
Count of Missions	-0.07	(0.04)	-0.07	(0.04)	-0.07	(0.04)	-0.07	(0.04)	-0.07	(-0.17,-0.02)
Change in Level	-0.05	(0.01)	-0.05	(0.01)	-0.05	(0.01)	-0.05	(0.01)	-0.05	(-0.08,-0.02)
Cumulative Payment	0.05	(0.02)	0.05	(0.02)	0.05	(0.02)	0.03	(0.02)	0.05	( 0.00, 0.10)
Count of Stayed Ties	-0.06	(0.02)								
Cumulative Count of Defected Ties	0.08	(0.02)								
Count of Stayed Strong Ties			-0.13	(0.10)	-0.13	(0.10)	-0.10	(0.10)	-1.31	(-2.52,-0.65)
Count of Stayed Weak Ties			-0.06	(0.02)	-0.07	(0.02)	-	-	-0.04	(-0.09,-0.01)
Cum. Count of Defected Strong Ties			0.18	(0.18)	0.16	(0.19)	0.25	(0.19)	1.62	( 0.43, 3.59)
Cum. Count of Defected Weak Ties			0.06	(0.03)	0.07	(0.27)	0.02	(0.25)	0.05	( 0.01, 0.11)
Log likelihood	-371.30		-369.33		-369.51		-374.92		-351.45	
AIC	762.60		762.66		763.03		771.84		724.90	

\* Standard deviation; \*\* 95% confidence intervals

In Table 2.6 we see no flip sign. Most of the estimates are consistently significant. But we do observe some variation in the scale of estimates. This is inevitable due to the small sample size of this research. The robustness test indicates that the impact of truncation, though unavoidable, is marginal on the estimation of our models.

**Table 2.6**  
**Estimation Results of Robustness Test**

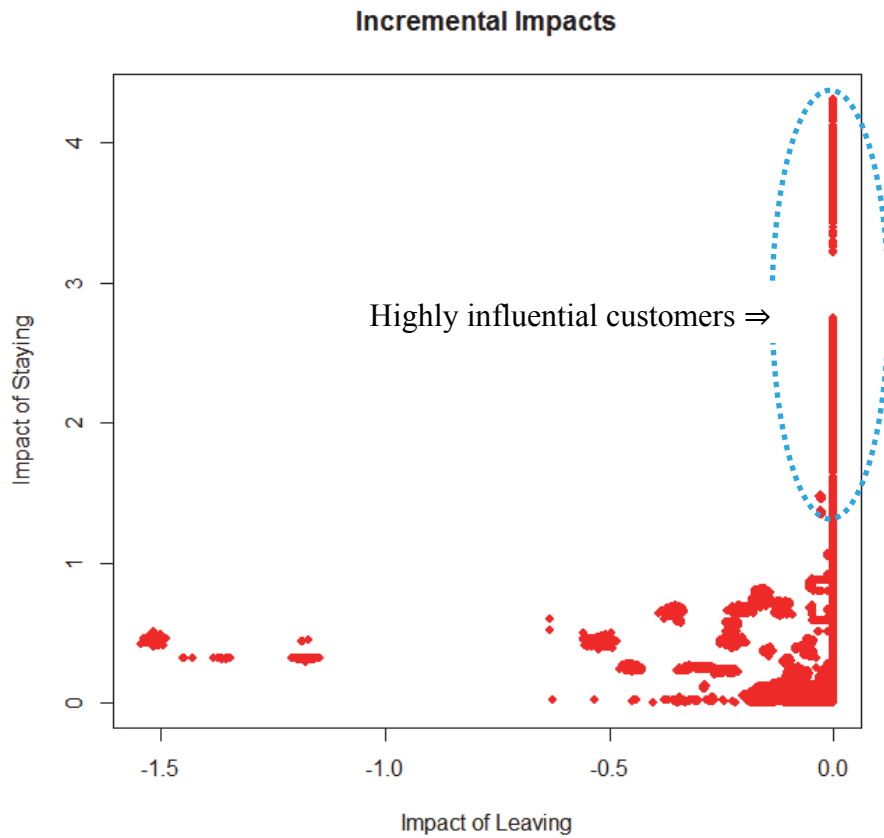
Data Set Used		T0	T1	T2	TA
Parameter Estimates	Intercept	-8.50	-11.01 *	-15.11 *	-9.36 *
	Partner Similarity	4.06	-0.59	1.68	-1.85
	Online Time Overlap	7.84 *	6.14 *	9.53 *	12.46 *
	Cumulative Count of Interactions	-0.45	-3.05	-2.39	-5.14
	Recency of Interaction	-4.65 *	-6.80 *	-4.79 *	-6.16 *
	Intercept	-6.87 *	-4.32 *	-5.44 *	-4.51 *
	Count of Guilds Attended	-0.42 *	-0.21 *	-0.12	-0.21 *
	Tenure in the Game	0.12 *	0.10 *	0.19 *	0.09 *
	ln(Online Time)	-0.21	-0.40	-0.23	-0.01
	ln(Cumulative Online Time)	0.43 *	0.14 *	0.06 *	0.18 *
	Count of Missions	-0.04	-0.15 *	-0.75 *	-0.07 *
	Change in Level	-0.05 *	-0.06 *	-0.07 *	-0.05 *
	Cumulative Payments	0.06	0.03 *	0.06 *	0.05 *
	Count of Stayed Strong Ties (SS)	-3.85 *	-2.10 *	-3.56 *	-1.31 *
	Count of Stayed Weak Ties (SW)	-0.05 *	-0.02 *	-0.06 *	-0.04 *
	Cum. Count of Defected Strong Ties (CDS)	3.07 *	2.40 *	4.44 *	1.62 *
	Cum. Count of Defected Weak Ties (CDW)	0.26 *	0.09 *	0.05 *	0.06 *
Number of Customers in the Data Set		70	147	103	181
Number of Dyads in the Data Set		141	415	701	1257

\* significant at 0.05 level.

### *Influential Customer Identification and Social Network Value*

With all the estimated parameters, we can then calculate each customer's incremental impact on other customers' defection decisions (equation (2.4) - (2.7)). This incremental impact can be used to identify influential customers. We illustrate each customer's total incremental impact in Figure 2.3.

**Figure 2.3**  
**Incremental Impacts of Customers' Defection Decisions**

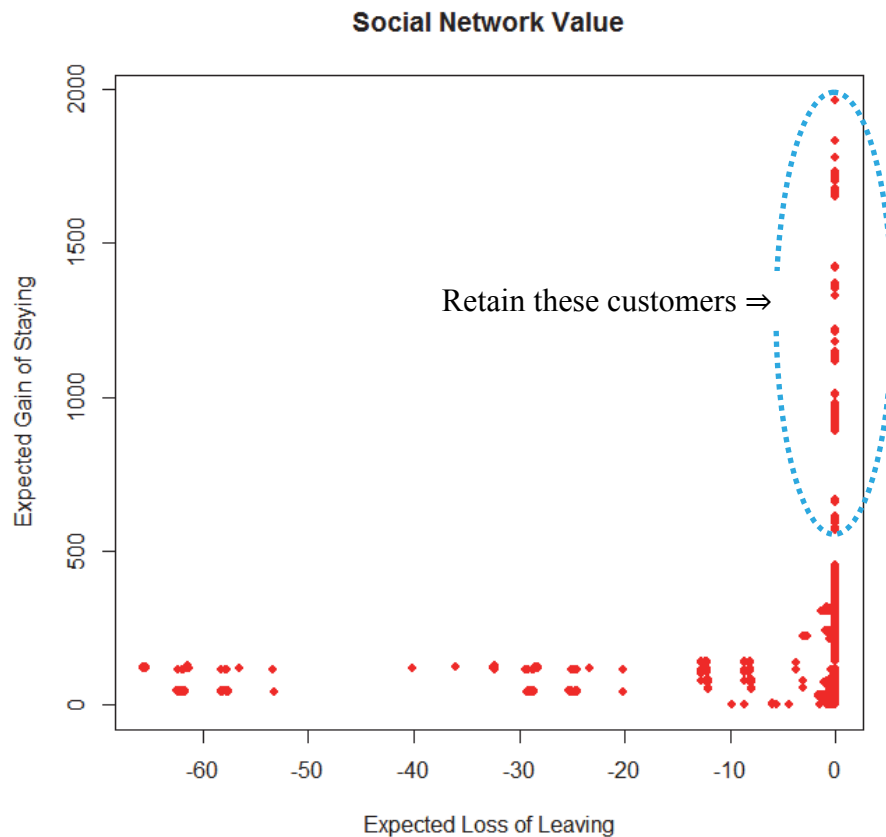


We can use equation (2.9) and customers' purchasing records to calculate each customer's social network value in each week. We can also calculate customers' total social network value during their tenures.

Note that there are two types of social network value measures here: (1) if a customer chooses to stay, she will have a *positive* network value (a potential lift on the expected value

contribution made by customers under her influence); (2) if a customer decides to leave, she will have a negative network value (a potential loss in the expected value contribution). We plot customers' total social network values in Figure 2.4.

**Figure 2.4**  
**Measuring Customers' Social Network Values**



It is evident in Figure 2.3 and 2.4 that there exists a small group of influential customers. If they choose to stay, the customers under their influence are more likely to stay. When they decide to leave, they will impact other customers to leave too.

Our tie strength model tells us that we can identify these influential customers using their interaction history: the more frequently a customer interacting with others, the more time she spends with others, the more likely she has strong influence on others.

Also note that if a customer is actively interacting with large, closely-connected groups,



she is likely to be influential. This is because groups with high centrality (Wasserman and Faust 1994) have large  $\exp(\mathbf{X}_{jt}^{SE} \cdot \boldsymbol{\beta} + \mathbf{X}_{jt}^{SC} \cdot \boldsymbol{\gamma})$ , which enhances the individual incremental impact (see equation (2.4) – (2.7)).

### *LOO Cross-Validation*

“Leave-One-Out” (LOO) is one type of the bootstrap algorithm used to cross-validate the predictive power of models. This approach is particularly suitable for social network models. Customers in a social network usually have complex connections with others, which makes it difficult to divide the data into an estimation set and a validation set. Using LOO algorithm, however, we can loop across all the customers, at each step removing one customer and her social ties. We estimate the models with the data of all the remaining customers; then use the estimated models to predict the defection behaviors of this “left-out” customer. Using this approach we can test the predictive power of our models without interfering too much with the whole network. One big disadvantage of this approach is that LOO is very time-consuming. If we have  $N$  customers we need to estimate the model  $N$  times. Particularly, when we use MCMC algorithm, time becomes a serious issue.<sup>31</sup>

The results of LOO validation are given in Table 2.7 and Figure 2.5.

**Table 2.7**  
**Results of LOO Validation**

(1) Prediction with Model

		Predicted Defections		
		No	Yes	Total
Actual Defections	No	28 (46.67%)	32 (53.32%)	60 (100.00%)
	Yes	49 (40.50%)	72 (59.50%)	121 (100.00%)

<sup>31</sup> In our case, estimating one model on a Linux server takes about a week.

**Table 2.7 (Continued)**  
(2) Prediction without Model

		Predicted Defections		
		No	Yes	Total
Actual Defections	No	0 (0.00%)	60 (100.00%)	60 (100.00%)
	Yes	0 (0.00%)	121 (100.00%)	121 (100.00%)

We validate our models at two levels. First, we test our models' capability of identifying churned customers. Second, for those churned customers who are correctly identified, we check how accurately our models can predict their defection time. As a benchmark, we use following "model-free" approach to predict customers' defection:

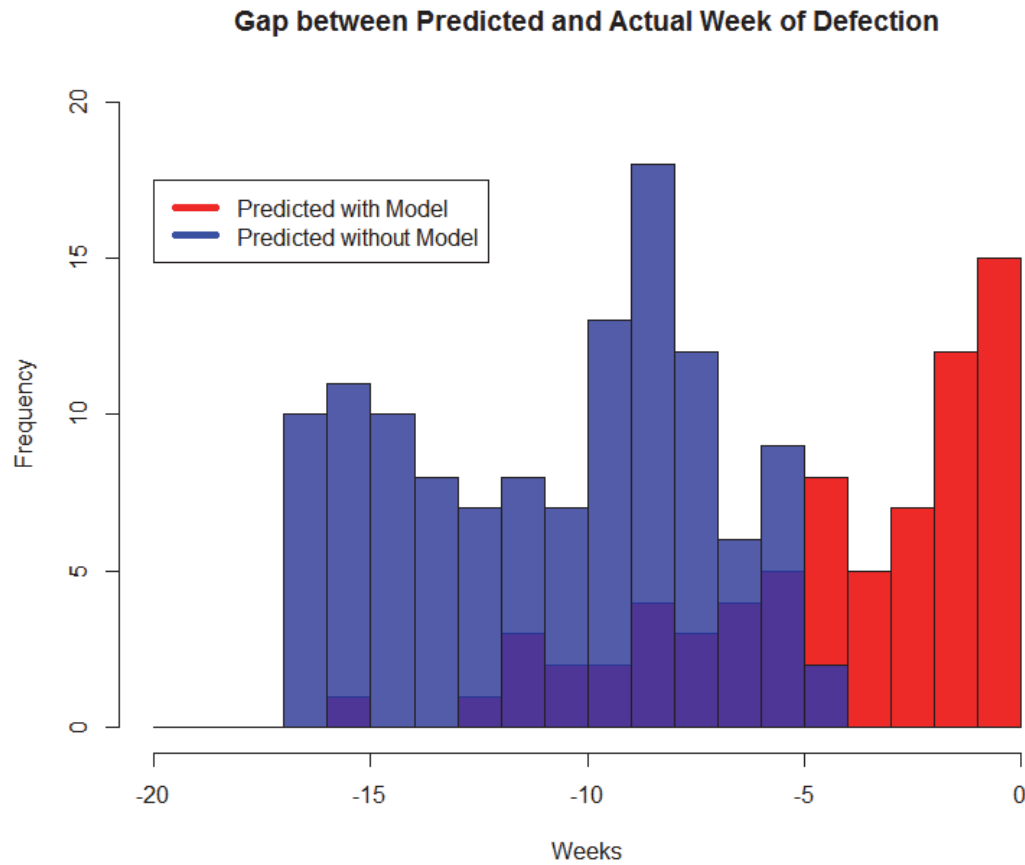
We assume that managers know the average defection rate in each week, which is 6% in our case. We assume that without a model, in each week managers simply randomly choose 6% customers and predict that they will defect within that week.

From Table 2.7 we can see that with our models we can correctly identify nearly 60% of the churned customers and 56.67% of the customers who decided to stay. "Model-free" approach captures all the defected customers by predicting that all customers have churned, thus resulting in a 100% misclassification of un-churned customers.

More importantly, if we take a closer look at these two approaches' predictions on defection time, we find that the modeling approach yields much higher accuracy than model-free approach. Modeling approach correctly predicts the defection time in 15 cases, and most of its predicting errors fall within five weeks. As a contrast, model-free approach makes no correct prediction on defection time. Most of its prediction errors are larger than five weeks (Figure 2.5).

**Figure 2.5**

**Model Can Predict Defection Time with Higher Accuracy**



Accurately predicting defection time is critical for managers. Managers want to identify potential churned customers before hand, and persuade them to stay. Premature identifications of churned customers could lead to suboptimal allocation of resources: time and money are spent on customers with little intention to defect. Figure 2.5 illustrates that using our models can help managers improve their allocation of resources to retain customers.

*Policy Simulation*

To demonstrate the importance of identifying and retaining influential customers, we set up two policy simulations. In the first simulation we assume that company manages to persuade the top 25% of their most influential customers to stay until the end of the game. In second

simulation company retains bottom 25% of their least influential customers until the end. The simulation results are listed in Figure 2.6.

**Figure 2.6**  
**Retention of Influential Customers Can Largely Increase Revenues**

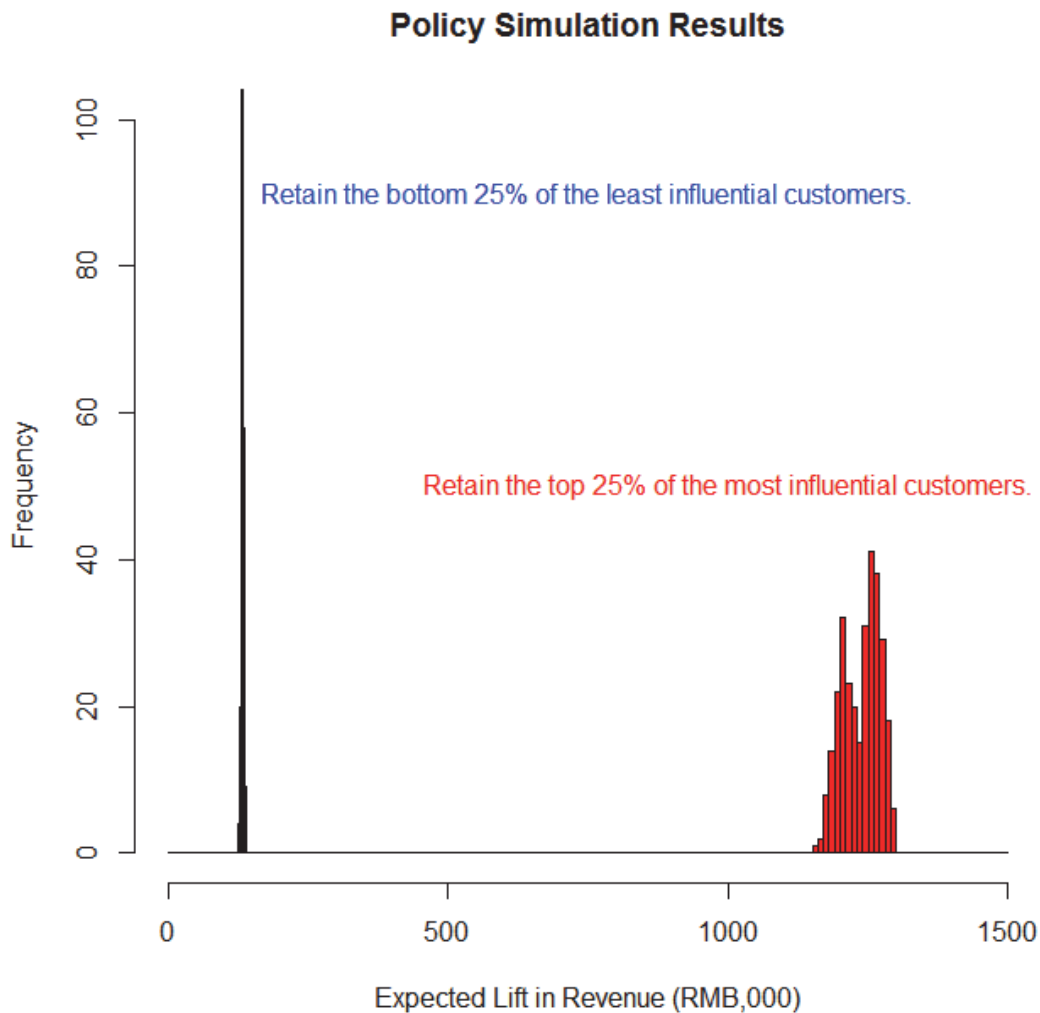


Figure 2.6 shows that both approaches increase revenues. And apparently retaining influential customers lead to much bigger increase in revenues. With this information, a company can decide how much it should spend to retain those influential customers. On average, retaining the top 25% most influential customers will generate 110,000 RMB (approx. 18,000 USD) more revenues than retaining the bottom 25% of the least influential customers.

## 2.7 Discussion

To the best of our knowledge, this research is the first attempt to model customers' defection behaviors within a social network. We jointly estimate a tie strength model and a defection decision model. Using defection decision model we capture the social influence on customers' defection through strong ties and weak ties. Using tie strength model we can identify influential customers with the records of their social interactions.

We find that if a customer is frequently interacting with other customers she tends to have strong tie with those customers. We also find that social contagion influence through strong ties is significantly stronger than the one through weak ties. With defection decision model we can also estimate the incremental impact of a customer's defection decision on the customers under her influence. With this measure we are able to identify influential customers in the network.

In this research we also propose a new measure of social network value, which is based on a customer's incremental impact on other customers' propensity to stay. Combining incremental impact measure with customers' purchasing records we calculate each customer's social network value, which is the lift on expected monetary contribution by other customers under influence. This social network value, together with customers own purchase (intrinsic value), can help a company more effectively allocate its CRM resources. We demonstrate the importance of correctly identifying and retaining influential customers with two policy simulations.

Our research is not without limitations. First, due to the small dataset we have (181 customers, 16 weeks) we cannot afford a more sophisticated model structure. Therefore, in this research we ignore customer heterogeneity and panel data characteristics. We use pooled

estimation. We also ignore the time-diminishing effects in customer's interactions and influence.<sup>32</sup>

Second, our data set suffers from left truncation issue. Even though in our robustness test section we have demonstrated that left truncation has only a marginal impact, it still biases our estimation results.

Third, in this research we only capture the direct (first order) social influence. We ignore the influence on "friend's friends". Therefore this conservative measure indicates the lower boundary of a customer's true social influence. We might underestimate the true value of a customer's social influence.

Forth, we directly use customers' purchasing records to calculate the social network value. We do not analyze what factors influence a customer's purchases. It could be very interesting to analyze customers' purchasing behaviors in the same framework as well.

Based on aforementioned limitations, our framework can be extended in the following four directions:

First, with richer dataset, we can introduce more complex structure to the tie strength model and the defection decision model, to addresses the specific requirements from panel data.

Second, it can be interesting to analyze the influence on friend's friend and see how the social influence gets diminished or enhanced. This will require more sophisticated network analysis techniques.

Third, we can integrate a purchasing model into our framework. Therefore we can have better understanding of not only the impact of social influence on customers' defection decisions, but also the impact on their purchasing decisions. With this knowledge we will be able to get a

---

<sup>32</sup> One finding from our prior research shows that customers have short memory about their past purchases, but long memory about their past interactions (for details, see Chapter 1). Hence in this research we do not consider the time-diminishing effect.

more comprehensive and accurate measure of customer's social network values.

Fourth, to make a full use of the information we can model tie strength with a continuous distribution (instead of a binary state). Accordingly, we can generate a series of expected values from this distribution and use them as social contagion covariates in the defection model. Therefore we will be able to have better understanding of the impact of social influence on customers' defection decisions.

## Appendix

The MCMC algorithm to jointly estimate tie strength model and defection decision model

*Generate  $\alpha$*

At each iteration,

- 1) Draw a new vector of  $\alpha$ :  $\alpha^{(k+1)} = \alpha^{(k)} + \Delta_\alpha$ , where  $\Delta_\alpha$  is drawn from a normal distribution:  
 $\Delta_\alpha \sim N(\mathbf{0}, \delta_\alpha \mathbf{I})$ ;  $\delta_\alpha$  is the size of random-walk step;
- 2) Draw latent tie strength states: For customer dyad  $(i, j)$  in week  $t$  we draw a random number from a normal distribution  $r_{(i,j),t} \sim N(\mathbf{X}_{(i,j),t}^{Hist'} \cdot \alpha, 1)$ . If  $r_{(i,j),t} > 0$  then customer  $i$  and  $j$  have a strong tie in week  $t$ :  $RS_{(i,j),t} = 1$ ; otherwise their tie strength is weak:  $RS_{(i,j),t} = 0$ ;
- 3) Update the four social contagion covariates for each customer  $i$  in each week  $t$ :  $SS_{it}$ ,  $SW_{it}$ ,  $CDS_{it}$ , and  $CDW_{it}$ ;
- 4) Calculate the likelihood  $L(\mathbf{RS}|\alpha^{(k+1)})$  using equation (2.3);
- 5) Accept  $\alpha^{(k+1)}$  with probability

$$\Pr(\text{Accept}) = \min \left\{ \frac{\left[ \exp \left( (\alpha^{(k+1)} - \alpha_0)' \Sigma_{\alpha_0} (\alpha^{(k+1)} - \alpha_0) \right) \right] \times L(\mathbf{RS}|\alpha^{(k+1)})}{\left[ \exp \left( (\alpha^{(k)} - \alpha_0)' \Sigma_{\alpha_0} (\alpha^{(k)} - \alpha_0) \right) \right] \times L(\mathbf{RS}|\alpha^{(k)})}, 1 \right\}. \quad (2.11)$$

where  $\alpha^0$  and  $\Sigma_{\alpha 0}$  are the hyperparameters of a normal prior. We use a diffuse prior:  $\alpha^0$  is a  $n_\alpha \times 1$  vector of zeros,  $\Sigma_{\alpha 0} = 100 \times I_{n_\alpha}$ , where  $n_\alpha = \dim(\alpha)$ .

*Generate  $\beta$*

1) Draw a new vector of  $\beta$ :  $\beta^{(k+1)} = \beta^{(k)} + \Delta_\beta$ , where  $\Delta_\beta$  is drawn from a normal distribution:

$\Delta_\beta \sim N(\mathbf{0}, \delta_\beta I)$ ;  $\delta_\beta$  is the length of step;

2) Calculate the likelihood using equation (2.3);

3) Accept  $\beta^{(k+1)}$  with probability

$$\begin{aligned} & \Pr(\text{Accept}) \\ &= \min \left\{ \frac{\left[ \exp \left( (\beta^{(k+1)} - \beta_0)' \Sigma_{\beta 0} (\beta^{(k+1)} - \beta_0) \right) \right] \times L(\mathbf{D} | \beta^{(k+1)})}{\left[ \exp \left( (\beta^{(k)} - \beta_0)' \Sigma_{\beta 0} (\beta^{(k)} - \beta_0) \right) \right] \times L(\mathbf{D} | \beta^{(k)})}, 1 \right\}. \end{aligned} \quad (2.12)$$

where  $\mathbf{D}$  is the vector of customers' defection decisions;  $\beta^0$  and  $\Sigma_{\beta 0}$  are the hyperparameters of a normal prior. We use a diffuse prior:  $\beta^0$  is a  $n_\beta \times 1$  vector of zeros,  $\Sigma_{\beta 0} = 100 \times I_{n_\beta}$ , where  $n_\beta = \dim(\beta)$ .

*Generate  $\gamma$*

1) Draw a new vector of  $\gamma$ :  $\gamma^{(k+1)} = \gamma^{(k)} + \Delta_\gamma$ , where  $\Delta_\gamma$  is drawn from a normal distribution:

$\Delta_\gamma \sim N(\mathbf{0}, \delta_\gamma I)$ ;  $\delta_\gamma$  is the length of step;

2) Calculate the likelihood using equation (2.3);

3) Accept  $\gamma^{(k+1)}$  with probability

$$\begin{aligned} & \Pr(\text{Accept}) \\ &= \min \left\{ \frac{\left[ \exp \left( (\gamma^{(k+1)} - \gamma_0)' \Sigma_{\gamma 0} (\gamma^{(k+1)} - \gamma_0) \right) \right] \times L(\mathbf{D} | \gamma^{(k+1)})}{\left[ \exp \left( (\gamma^{(k)} - \gamma_0)' \Sigma_{\gamma 0} (\gamma^{(k)} - \gamma_0) \right) \right] \times L(\mathbf{D} | \gamma^{(k)})}, 1 \right\}. \end{aligned} \quad (2.13)$$

where  $\mathbf{D}$  is the vector of customers' defection decisions;  $\gamma^0$  and  $\Sigma_{\gamma 0}$  are the hyperparameters of



a normal prior. We use a diffuse prior:  $\boldsymbol{\gamma}^0$  is a  $n_{\boldsymbol{\gamma}} \times 1$  vector of zeros,  $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}^0} = 100 \times \boldsymbol{I}_{n_{\boldsymbol{\gamma}}}$ , where  $n_{\boldsymbol{\gamma}} = \dim(\boldsymbol{\gamma})$ .

## REFERENCES

- Allenby, Greg M., Robert P. Leone, and Lichung Jen (1999), "A Dynamic Model of Purchase Timing with Application to Direct Marketing," *Journal of the American Statistical Association*, 94 (June), 365-374.
- Anderson, Eugene W. (1998), "Customer Satisfaction and Word of Mouth," *Journal of Service Research*, 1(1), 5-17.
- Aral, Sinan, Lev Muchnik, and Arun Sundararajan (2009), "Distinguishing Influence-Based Contagion from Homophily-Driven Diffusion in Dynamic Networks," In *Proceedings of the National Academy of Sciences*. 106.51: 21544-21549.
- Aral, Sinan, and Dylan Walker (2012), "Identifying Influential and Susceptible Members of Social Networks," *Science*, 337.6092 (2012): 337-341.
- Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson (2012), "The Diffusion of Microfinance," working paper.
- Bhatt, Rushi, Vineet Chaoji, and Rajesh Parekh (2010), "Predicting Product Adoption in Large-Scale Social Networks," In *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 2010.
- Billington, Jim (1996), "Five Keys to Keep Your Best Customers," *Harvard Management Update*, July.
- Bolton, Ruth N. (1998), "A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: The Role of Satisfaction," *Marketing Science*, 17(1), 45-65.
- Centola, Damon (2010), "The spread of behavior in an online social network experiment," *Science*, 329.5996: 1194-1197.
- Chandrashekar, Murali, Kristin Rotte, Stephen S. Tax, and Rajdeep Grewal (2007), "Satisfaction Strength and Customer Loyalty," *Journal of Marketing Research*, 44 (February), 153-163.
- Christakis, Nicholas A. and James H. Fowler (2007), "The Spread of Obesity in a Large Social Network over 32 Years," *The New England Journal of Medicine*, 357:370-9.
- Domingos, Pedro, and Matt Richardson (2001), "Mining the Network Value of Customers," In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2001.
- Fader, Peter S. and Bruce G.S. Hardie (2009), "Probability Models for Customer-Base Analysis," *Journal of Interactive Marketing*, 23, 61-69.

- Fowler, James H. and Nicholas A Christakis (2008), "Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis Over 20 Years in the Framingham Heart Study," *BMJ*, 337:a2338.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, Donald B. Rubin (2004), *Bayesian Data Analysis*, Chapman & Hall/CRC.
- Gupta, Sunil (1991), "Stochastic Models of Interpurchase Time with Time-Dependent Covariates," *Journal of Marketing Research*, 28 (February), 1-15.
- Gupta, Sunil, Carl F. Mela, and Jose M. Vidal-Sanz (2009), "The Value of a "Free" Customer," *Working Paper*.
- Iacobucci, Dawn, and Nigel Hopkins (1992), "Modeling Dyadic Interactions and Networks in marketing," *Journal of Marketing Research*, 29 (February), 5-17.
- Iyengar, Raghuram, Christophe Van den Bulte, and Thomas W. Valente (2011), "Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science*, 30(2), 195-212.
- Jones, Thomas O. and W. Earl Sasser, Jr. (1995), "Why Satisfied Customer Defect," *Harvard Business Review*, November-December.
- Katona, Zsolt, Peter Pal Zubcsek, and Miklos Sarvary (2011), "Network Effects and Personal Influences: The Diffusion of an Online Social Network," *Journal of Marketing Research*, 48 (June), 425-443.
- Lehonvirta, Vili (2009), "Virtual Item Sales as a Revenue Model: Identifying Attributes that Drive Purchase Decisions," *Electronic Commerce Research*, 9, 97-113.
- Manski, Charles F. (1993), "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 60(3), 531-542.
- Nair, Harikesh S., Puneet Manchanda, and Tulikaa Bhatia (2010), "Asymmetric Social Interactions in Physician Prescription Behavior: The Role of Opinion Leaders," *Journal of Marketing Research*, 47(October), 883-895.
- Neslin, Scott A., Sunil Gupta, Wagner Kamakura, Junxiang Lu, and Charlotte H. Mason (2006), "Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models," *Journal of Marketing Research*, 43 (May), 204-211.
- Nitzan, Irit and Barak Libai (2011), "Social Effects on Customer Retention," *Journal of Marketing*, 75 (November), 24-38
- Reichheld, Frederick F. (1996), "Learning from Customer Defections," *Harvard Business Review*, March-April.

Reinartz, Werner J. and V. Kumar (2000), "On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing," *Journal of Marketing*, 64 (October), 17-35.

Reinartz, Werner J. and V. Kumar (2003), "The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration," *Journal of Marketing*, 67 (January), 77-99.

Reinartz, Werner J. and Rajkumar Venkatesan (2008), "Decision Models for Customer Relationship Management (CRM)," *Handbook of Marketing Decision Models*, Springer Science + Business Media, LLC.

Schmittlein, David C., Donald G. Morrison, Richard Colombo (1987), "Counting Your Customers: Who Are They and What Will They Do Next?" *Management Science*, 33(January), 1-24.

Schmittlein, David C. and Robert A. Peterson (1994), "Customer Base Analysis: An Industrial Purchase Process Application," *Marketing Science*, 13(1), 41-67.

Seetharaman, P. B. and Pradeep K. Chintagunta (2003), "The Proportional Hazard Model for Purchase Timing," *Journal of Business & Economic Statistics*, 21:3, 368-382.

Shankar, Venkatesh, Amy K. Smith, and Arvind Rangaswamy (2003), "Customer Satisfaction and Loyalty in Online and Offline Environments," *International Journal of Research in Marketing*, 20, 153-175.

Trusov, Michael, Anand Bodapati, and Randolph E. Bucklin (2010), "Determining Influential Users in Internet Social Networks," *Journal of Marketing Research*, 47 (August), 643-658.

Villanueva, Julian, Shijin Yoo, and Dominique M. Hanssens (2008), "The Impact of Marketing-Induced Versus Word-of-Mouth Customer Acquisition on Customer Equity Growth," *Journal of Marketing Research*, 45(February), 48-59.

Wasserman, Stanley and Katherine Faust (1994), *Social Network Analysis: Methods and Applications*, New York: Cambridge University Press.

Williams, Dimitri, Nick Yee, Scott E. Caplan (2008), "Who Plays, How Much, and Why? Debunking the Stereotypical Gamer Profile," *Journal of Computer-Mediated Communication*, 13(4), 993-1018.

Xiang, Rongjing, Jennifer Neville, and Monica Rogati (2010), "Modeling Relationship Strength in Online Social Networks," In *Proceedings of the 19th International Conference on World Wide Web*. ACM, 2010.

### **CHAPTER 3**

## **CUSTOMER RELATIONSHIP MANAGEMENT AND SOCIAL NETWORK ANALYSIS: POSSIBLE SYNERGY OPPORTUNITIES**

### ***Abstract***

This paper discusses the promising research opportunities in integrating social network analysis (SNA) components into customer relationship management (CRM). The aim is to enable firms to manage their customers as a network and leverage the power of social influence among customers to enhance customer-firm relationships. This paper briefly reviews the four critical aspects of CRM: acquisition, retention, growth, and firm-customer relationship dynamics. Within each aspect the discussion focuses on the possible impact of social network components on CRM models, and how to combine CRM and SNA in modeling efforts. It can be foreseen that, with the increasing availability of CRM+SNA data, more research will be carried out in these interesting areas.

**Key Words:** customer relationship management, social network analysis

### **3.1 Introduction**

The main purpose of this dissertation is to explore a way to leverage the power of social influence among customers within a social network, and enhance their relationships with the company. In Chapter 1 the discussion concentrates on customers' influence on each other's purchasing behaviors through strong ties and weak ties in a social network. In Chapter 2 the investigation focuses on modeling the impact of social influence on customers' defection decisions.

These two projects have revealed enormous research potential in the intersection of customer relationship management (CRM) and social network analysis (SNA). In classical customer relationship management research customers are treated independently. This practice might no longer be appropriate for customers connected within a social network. More sophisticated methodology is needed to model the impact of social influence on customers' behaviors and their relationships with companies. On the other hand, social network analysis research, even though undergoing fast growth in marketing area, rarely crosses path with CRM research. Therefore, as Rust and Chung (2006) point out, investigating how to manage customers as a network is a promising direction for future CRM research. And that is the topic of the first two chapters of this dissertation.

As a natural extension of Chapters One and Two, this chapter searches for and identifies possible synergy opportunities between customer relationship management and social network analysis. Beginning with a brief review of the recent developments in both CRM and SNA, this chapter discusses the possibilities to integrate social network components into the four key components of CRM: customer acquisition, customer retention, customer growth, and firm-customer relationship dynamics.

In “acquisition” section, the discussion focuses on the use of social referral and social media to acquire customers. In “retention” section, the discussion concentrates on the impact of social influence on customers’ satisfaction, churn, and reaction to service failure/recovery. In “growth” section, the discussion covers the possibility to promote new products, or cross-sell sequentially-purchased products to the incumbent customers in a social network. In “relationship dynamics” section, the use of hidden Markov model (HMM) is reviewed, in its role to capture the latent, dynamic relationship between a firm and its customers. Then the discussion moves on to the hope and difficulty to consider social effects in modeling firm-customer relationship dynamics.

### **3.2 CRM and SNA – A Brief Review**

#### *Customer Relationship Management*

The idea of CRM can be traced back to the 1980’s and the term of “CRM” appeared in the 1990’s (Payne and Frow 2005). Since then, despite some implementation difficulties, CRM’s popularity has been growing fast among top executives. For instance, a 2008 survey by Gartner Inc. showed that 75% of senior managers planned to invest on CRM in the coming year (Mertz 2009).

At the same time, research on CRM also gained tremendous momentum. An important mission of CRM research is to find a way to allocate resources, in a differentiated and systematic way, to customers with different economic values (Reinartz and Venkatesan 2008). Accordingly, researchers need an accurate measurement of customer life time value (CLV), which is “the present value of future profit generated from a customer over his/her life of business with the firm” (Gupta and Lehmann 2008, page 256).

Furthermore, the sum of individual customers' CLV is defined as a firm's current customer equity (or assets) (Reinartz and Venkatesan 2008). Some researchers propose to use a firm's customer equity as a link between its marketing activities and its shareholder value (e.g., Hogan et al. 2002, Gupta and Zeithaml 2006).

In order to accurately measure CLV, researchers need to have a thorough understanding of various relationship phases between a company and its customers. According to Reinartz and Venkatesan (2008), there are five key components in CRM: acquisition, retention, growth, winning back lost customers, and firm actions. Researchers have developed various models to address the issues in these five key components of CRM. In order to facilitate the following discussion on CRM+SNA, "winning back lost customers" is considered as a specific type of retention effort. As for firm actions component, the discussion focuses on the relationship dynamics between a firm and its customers, due to its CRM activities. Some representative CRM papers are categorized in Table 3.1.

#### *Social Network Analysis in Marketing*

Compared with CRM, social network analysis has a much longer history. The first social network research was published in the 1930's (Moreno 1934, in which the concept of "sociometry" and "sociogram" were introduced). After 40 years, contemporary social network analysis took shape between 1960's and 1970's (Scott and Carrington 2011). After that this research area underwent slow growth from 1970's to 1990's (Carrington, Scott and Wasserman 2005). Since then social network research has been expanding rapidly, due to the development of mathematical tools to model networks (Alba 1982). These days social network analysis has been applied to areas such as sociology, psychology, physics, education, communication, politics, technologies, economics, and management (Scott and Carrington 2011).



**Table 3.1**  
**Research on CRM**

(a) Acquisition

Topic	Paper	Model/Method	Findings	Extension Opportunities
Direct marketing / Using profit maximization to decide the cutoff value for target customers	Bult and Wansbeek (1995)	RFM model	Profit maximization can help researchers understand customer response curves better	Purchase amount, frequency, and features of direct marketing can be considered
Direct marketing / Investigating the performance of model directly predicting customer's discrete response	Bodapati and Gupta (2004)	Discretized response scoring model	When sample size is large and model is mis-specified, modeling discretized responses outperforms modeling continuous responses	Model can be extended to predict customers' purchasing intention of new product concept
Customer profiling / Discussing the use of massively categorical variable (e.g. zip code) as explanatory covariate	Steenburgh et al. (2003)	Hierarchical Bayes variance component model	The model performs well to estimate parameters. Massively categorical variables can be used to replicate customer information	Semiparametric and nonparametric techniques can be used
Acquisition channel / Comparing customer acquisition through various channels, in terms of loyalty and cross-selling opportunities	Verhoef et al. (2005)	Probit model	Customers acquired through website channels are more loyal than the ones acquired through direct mail, radio, and TV	Message contents can be considered. Analysis over long term is needed
Acquisition channel / Comparing the short-term and long-term value of customers acquired through WOM channel and marketing channels	Villanueva et al. (2008)	VAR model	Customers acquired through marketing-induced channels have higher short-term value, while customers acquired through WOM channel have higher long-term value	Dynamic interactions among acquisition channels can be analyzed. WOM creation can be modeled
Customer referral / Modeling customer referral value (CRV)	Kumar et al. (2010)	Bayesian Tobit model	Managers should measure CLV and CRV separately. Understanding the drivers for CRV and targeting referral campaign at customers in low-CRV segment can financially benefit companies	This research concentrates on the extrinsic motivation (incentive), customers' intrinsic motivation to make referrals can be analyzed

**Table 3.1 (Continued)**

Topic	Paper	Model/Method	Findings	Extension Opportunities
Customer referral / Investigating to what extent customers acquired through referral programs are more valuable than others	Schmitt et al. (2011)	Linear regression and proportional hazard model	In financial service sector, compared with non-referral customers, referred customers have higher retention rate and contribute higher margin (but eroding with time)	Actual social mechanism (such as dyadic connections) can be considered in the model
(b) Retention				
Topic	Paper	Model/Method	Findings	Extension Opportunities
Customer status / Using customer purchase history to predict customer activeness status	Schmittlein et al. (1987)	NBD/Pareto model	Developed first mathematically tractable and realistic model to predict a customer's probability to be alive	Time-varying covariates can be integrated into the model. Customer heterogeneity can be modeled with hierarchical structure
Interpurchase timing / Predicting the change in customer purchase behaviors	Allenby et al. (1999)	Hierarchical Bayesian model	Model can be used to identify customers in different active states. Companies can allocate resources accordingly	More covariates can be included in the random-effect model
Satisfaction and retention / Capturing the relationship between customer satisfaction and retention	Bolton (1998)	Proportional hazard model	Customers' satisfaction is positively related with the duration of their relationship with service providers. The strength of "satisfaction-duration" relationship depends on the length of prior experience with service providers	Time-varying marketing covariates can be considered. Impact of customer satisfaction on their behaviors can be analyzed
Satisfaction and retention / Analyzing the relationship between customer retention and satisfaction and other drivers	Gustafsson et al. (2005)	Linear regression model (aggregate level)	Customer satisfaction has negative impact on customer churn, while affective commitment has no significant effect	Better measures of satisfaction and affective commitment can be explored. Limitation on data collection (through survey) needs to be addressed
Acquisition and retention / Investigating the way to balance resource allocation on customer acquisition and retention	Reinartz et al. (2005)	Acquisition/retention two stage modeling	Suboptimal resource allocation for retention has greater impact on long-term profitability than suboptimal allocation for acquisition. Highly interpersonal and interactive communication benefits a company if it's initiated by the company	Competition factor can be considered. Marketing expenditure in B2C setting needs more attention.

**Table 3.1 (Continued)**

Topic	Paper	Model/Method	Findings	Extension Opportunities
Defection prediction / Analyzing the performance of various models predicting customer defection	Neslin et al. (2006)	Meta-analysis	In contractual situation probit / logit models perform better than other models	Attention can be paid to missing-value techniques, machine learning, and nonparametric methods. Dynamic procedure can be used
Win back lost customers / Determining the optimal pricing strategy to re-capture lost customers	Thomas et al. (2004)	Probit model / regression model	Offering both low re-acquisition price and retention price is an optimal strategy. If price offered in re-established relationship is lower than the first relationship, customers tend to have longer tenure	Impact of pricing on demand and pricing elasticity can be considered
Service recovery / Analyzing customer satisfaction with service failure / recovery encounters	Smith et al. (1999)	Linear regression model	Customers want to receive recovery resources that match the type and magnitude of the service failures they experience	Customer heterogeneity in their response to service failures can be considered. Conjoint approach can be used to design recovery strategies
(c) Growth				
Topic	Paper	Model/Method	Findings	Extension Opportunities
Cross-selling / Predicting the next product an individual customer is most likely to purchase	Knott et al. (2002)	"Next product to buy" model	Neural net models perform better than statistical models in terms of prediction accuracy	The model can be combined with CLV models. The model can also be extended to be an optimal product targeting model
Cross-selling / Using customers' past purchase history across multiple categories to predict the purchase of a new product	Kamakura et al. (2004)	Multivariate split-hazard model	The model performs well in predicting the adoption of existing and new products	Marketing factors can be considered. The model can be extended to address repeat purchase. The model can be embedded in an optimal product framework
Cross-selling in B2B / Jointly modeling product choice and purchase timing	Kumar et al. (2008)	Purchase timing: log-logistic model Category choice: probit model	Customer-focused campaign (promoting products only when customers are expected to purchase) can generate more profits than conventional sales campaigns	The model might not apply to B2C setting. Purchase quantity can be modeled. Endogeneity issue might need to be addressed

**Table 3.1 (Continued)**

Topic	Paper	Model/Method	Findings	Extension Opportunities
CLV estimation / Estimating CLV in a non-contractual setting	Reinartz and Kumar (2003)	NBD/Pareto + proportional hazard model	Proposed framework outperforms traditional RFM and other CLV frameworks. Customers with intermediate purchase frequency tend to have long relationship duration	Two-step procedure can be integrated into one framework. Satisfaction and loyalty factors can be considered. Marketing mix can also be considered
(d) Relationship Dynamics				
Topic	Paper	Model/Method	Findings	Extension Opportunities
Alumni relationship / Using transaction data to reveal the dynamic, latent firm-customer relationships	Netzer et al. (2008)	Hidden Markov Model (HMM)	The model preforms probabilistically well to segment customers into various relationship states	Multivariate outcome of relationships can be modeled. Nonstationary HMMs can be further analyzed. Surveyed attitudinal data can be integrated into the model
Pharmaceutical detailing / Modeling dynamic allocation of marketing resources based on the short-term and long-term effectiveness of marketing activities	Montoya et al. (2010)	HMM + dynamic programming	The framework provides important implications for dynamically managing customer relationship and maximizing long-term profitability	Endogeneity issue, forward-looking behavior, and social interaction can be considered

The application of social network analysis in marketing can be traced back to B2B research in the 1970's (Spekman 1996). In B2C area, research topics vary from new product adoption to social media and viral marketing. Table 3.2 lists some representative application of social network analysis in the marketing area.

**Table 3.2**  
**Social Network Analysis (SNA) Research in Marketing**

(a) Network Structure

Topic	Paper	Model/Method	Findings	Extension Opportunities
Interaction and relationship / Modeling the probability of each dyad connecting with certain relationship strength	Iacobucci and Hopkins (1992)	Hierarchical log-linear model	One of the first attempts to model relationship strength in a sequential interaction setting	The model involves large number of parameters, which might be difficult to estimate over a large network
Interaction and relationship / Simultaneously modeling multiple relationship of different types	Ansari et al. (2011)	Hierarchical Bayesian model	Heterogeneity, latent space, and relationship correlation are important components to recover the relationship structure. Researchers can use one type of relationship to predict the existence of another type of relationship	Relationship dynamics can be incorporated into the model. Bayesian nonparametric methods can be considered
Propagation on network / Investigating the impact of size and structure of local network around a node on the diffusion of products from this node	Yoganarasimhan (2012)	Descriptive dynamic model	The size and structure of local network around a node have significant impact on the diffusion of product from this node. There is temporal variation in this impact	The network can be extended to composite network, with data collected from multiple networks
Adoption / Analyzing how diffusion curve is affected by network	Dover et al. (2012)	Network-based growth model	Adopter network's degree distribution significantly affects the contagion properties of dissemination process	Research can be advanced to model aggregate patterns (adoption curve) on the heterogeneous individual behaviors
Acquisition channel / Comparing customer acquisition through various channels, in terms of loyalty and cross-selling opportunities	Verhoef et al. (2005)	Probit model	Customers acquired through website channels are more loyal than the ones acquired through direct mail, radio, and TV	Message contents can be considered. Analysis over long term is needed

**Table 3.2 (Continued)****(b) Word-of-Mouth**

Topic	Paper	Model/Method	Findings	Extension Opportunities
Customer satisfaction / Checking whether dissatisfied customers are more or less likely to engage in WOM	Anderson (1998)	Linear regression model	An asymmetric U-shape relationship exists between customer satisfaction and WOM. Extremely satisfied customers are most likely engaged in WOM	More flexible techniques can be used to estimate the relationship. Other attitudinal, behavioral and financial aspects related with customer satisfaction can be modeled
Drivers of WOM / Analyzing the psychological drivers of immediate and ongoing WOM	Berger and Schwartz (2011)	Poisson log-normal mixed model	More publicly visible products receive more immediate, ongoing and overall WOM. Promotional giveaways might promote WOM	Other product features might be considered. The impact of WOM on product diffusion can be analyzed. The impact of online and offline WOM can be compared
Online conversation / Investigating how to measure WOM from customer online conversation	Godes and Mayzlin (2004)	Linear regression model	Online conversation provides less costly but efficient way to measure WOM. Dispersion of conversation is valuable information for measuring WOM. Measurement should be carried out early in product's life cycle	Further investigation is needed to analyze why dispersion is important for the measurement of WOM. Casualty between WOM and sales needs to be established
Quality signal / Identifying and measuring contagious WOM	Nam et al. (2010)	Discrete time proportional hazard model	The effect of contagious WOM is asymmetric. Effect of negative WOM is stronger than the one of positive WOM	More service quality factors and customer heterogeneity can be considered in modeling

**(c) Opinion Leader**

Topic	Paper	Model/Method	Findings	Extension Opportunities
New product adoption / Measuring the impact of social interaction and peer effect in a setting of physician prescription	Nair et al. (2010)	Linear model	Asymmetric peer effects exist between opinion leaders and followers. Opinion leaders are more responsive to marketing activities	Opinion leaders and followers' heterogeneity, network characteristics, and marketing activities can be considered
New product adoption / Analyzing the impact of opinion leader and social contagion on new product adoption	Iyengar et al. (2011)	Linear model	The correlation between sociometric and self-reported measure of opinion leadership is weak. Social contagion exists through network ties and its effect depends on recipients' perception of opinion leadership	The role of product usage in contagion processes can be investigated

**Table 3.2 (Continued)**

(d) Social Contagion / Influence

Topic	Paper	Model/Method	Findings	Extension Opportunities
New product adoption / Identifying and measuring social contagion effect	Manchanda et al. (2008)	Hierarchical Bayesian model	Marketing communication has more significant impact on early adoption, while social contagion plays more and more important role with time	Time-varying marketing and contagion covariates can be considered. Asymmetric contagion effects can be investigated
New product adoption / Investigating the impact of network structure, influencer characteristics, and adoption characteristics on adoption process within a social network	Katona et al. (2011)	Binary choice model	Significant degree effect and clustering effect exist in adoption process. The characteristics of influencers and adopters also have significant impact on adoption	Endogeneity issue can be addressed. The assumption of homogeneous network can be relaxed
New product adoption / Measuring the degree of social contagion in product diffusion	Du and Kamakura (2011)	Discrete-time proportional hazard model	Evidence of social contagion in the diffusion of new consumer packaged products is found, if spatial and temporal heterogeneity is allowed in modeling	The mechanism of social contagion can be investigated
New product adoption / Using social interaction data to improve the forecast of new product adoption	Toubia et al. (2011)	Diffusion model	Using both social interaction data and penetration data can improve diffusion forecast	The impacts of WOM, network structure, relationship types, and marketing mix can be considered in the model
Formal and informal influence / Identifying formal and informal social influence in new technology adoption	Tucker (2008)	Binary discrete choice model	Individual-level causal network externality is identified. The influential is not necessarily the one with formal authority. It is important to target the people at the center of communication network	The effect of WOM needs to be investigated
Influential customers / Identifying customers with strong influence in a social network	Trusov et al. (2010)	Poisson regression model	Relatively fewer friends have significant influence on others' behaviors. Customer profile cannot help identify influential customers	Hidden Markov model can be employed to investigate the dynamic aspect. Multiple sites, offline social interactions can be analyzed. Influential customers' response to marketing actions can be modeled

**Table 3.2 (Continued)****(e) Social Contagion / Influence**

Topic	Paper	Model/Method	Findings	Extension Opportunities
Product attribute preference / Investigating how peers influence each other's preference for product attributes in a social network	Narayan et al. (2012)	Hierarchical Bayesian model	Customers update their preference for product attributes in a Bayesian manner, under peer influence from other customers	Peer effects on related purchase decisions can be analyzed. Both observational learning and information sharing can be investigated
Experience attributes / Analyzing how customers resolve uncertainty about products through social learning	Lee and Bell (2013)	Bayesian learning model	Local social learning can reduce uncertainty and lead to product trial. Local social capital lifts the efficiency of social learning	Other temporal and spatial effects can be controlled to establish the causal relationship. Other social contagion mechanism can be investigated

**(f) User-generated Contents**

Topic	Paper	Model/Method	Findings	Extension Opportunities
Social capital / Analyzing the impact of various factors on the release time of user-generated contents	Mallapragada et al. (2012)	Split hazard model	Open source project founders' social capital is a more important factor than product characteristics and the interplay between developers and end users. The existence of forums can accelerate product release	Contexts other than user-generated contents can be analyzed. The impact of type and quality of communication can be investigated
Collaborative contents / Investigating the impact of the number and embeddedness of contributors on the user value of user-generated contents	Ransbotham (2012)	Social network analysis and hierarchical linear latent model	A curvilinear relationship exists between the number of content contributors and the value of contents. Content contributors are not equal in value. The number and network of contributors have stronger impact at early stage	The transfer of specific contents can be examined. Endogeneity issues require further investigation
Social tie formation / Investigating the relationship between online users' content generation and their social ties	Shriver et al. (2013)	Linear model	Social ties can help content generation on a social website	Tie strength can be considered. The link between user-generated contents and advertising revenue can be analyzed. Social influence mechanism can be investigated



**Table 3.2 (Continued)**

(g) Social Media and Viral Marketing

Topic	Paper	Model/Method	Findings	Extension Opportunities
Viral product design / Examining the effectiveness of viral features on the generation of WOM peer influence	Aral and Walker (2011)	Hazard model	Passive broadcast viral features are more effective than active-personalized viral features in generating peer influence and contagion. Active-personalized viral features are more effective in encouraging adoption and sustainable product use	Optimal viral marketing strategy design based on the effectiveness of viral features can be investigated
WOM seeding / Analyzing how WOM seeding program can create value by expanding customer base and accelerating purchase by current customers	Libai et al. (2013)	Agent-based simulation model	For similar brands most of the value from seeding program is created through market expansion. Seeding programs targeting most influential customers generate more social value through acceleration	Social influence other than WOM, such as network externalities, can be investigated. The direction of communication can be considered. Customer life time value can be integrated
Social media marketing strategy / Proposing and measuring the metric of customer influence value. Measuring the ROI of social media marketing strategy	Kumar et al. (2013)	Field experiment	The proposed method measures both tangible and intangible value of social media, which can help optimize social media campaigns	The method can be tested indifferent industry settings. Customer heterogeneity can be considered. Game theory framework can be integrated

(h) Social Network Value

Topic	Paper	Model/Method	Findings	Extension Opportunities
Value of buyers and sellers / Measuring customer value in a network setting	Gupta et al. (2009)	Structural model	Strong direct (buyer-to-buyer) and indirect (buyer-to-seller) network effect exist in the network. Price and advertising elasticity becomes lower when network effects are considered	Models can be built to predict future demand, expenses and customer value
Value of networked sellers / Analyzing the value created through seller connections in an online commerce website	Stephen and Toubia (2010)	Hierarchical Bayesian model	Connection among sellers can generate enormous economic value, due to customers' accessibility to marketplace. Sellers with highest accessibility benefit the most	Different types of online marketplaces need to be explored. Online stores' influence on each other can be considered. Dynamics across time might be investigated

### **3.3 Acquire New Customers through Social Network and Social Media**

The key to successful CRM is finding the “right” customers to serve. The right customers are the ones with high life time value. However, it is difficult for a firm to predict its prospect customers’ CLV before actually acquiring them, especially when information is limited.

If prospect customers’ profiles are known, binary choice models such as logit or probit models can be used to analyze what type of customers are more likely to be acquired (e.g., Bodapati and Gupta 2004, Reinartz et al. 2005). If customer profiles are not available, diffusion models (Bass 1969) can be used (e.g., Bass et al. 1994, Kim et al. 1995, Gupta et al. 2004, Libai et al. 2009).

Above approaches treat customers as they are independent from each other. If customers are connected in a social network, the information of social connections and interactions can be used in acquisition models. For instance, Toubia et al. (2009) use customers’ social interaction covariates in a diffusion model to predict the adoption of a new product.

These days, omnipresent social networks and social media provide companies good opportunities to profile their incumbent and prospect customers in terms of their social behaviors. This enables companies to acquire new customers through social channels such as social referral and social media marketing.

Referral system has caught researchers’ attention for a more than a decade. For example, Biyalogorsky et al. (2001) discuss the use of referral reward. Kumar et al. (2010) propose to use both CLV and CRV (customer referral value) to identify the most valuable customers in terms of their referral activities. The authors also point out

the importance of understanding the drivers behind customers' referral value.

### *Effectiveness of New Acquisition Channels*

Before modeling the acquisition process through social referral (word-of-mouth, WOM) and social media channels, it is necessary to compare the effectiveness of these two new acquisition channels and the traditional channels such as advertising. Villanueva et al. (2008) propose a method to do so. In their paper the authors use vector autoregression (VAR) model to compare the impact of WOM acquisition and marketing-induced acquisition, in terms of company's financial performance. They find that customers acquired through word of mouth are comparatively more loyal and more valuable than the customers acquired by marketing campaigns.

Following this approach, the effectiveness of customer acquisition through social media marketing can also be compared:

$$\begin{bmatrix} MKT_t \\ WOM_t \\ SOCIAL_t \\ VALUE_t \end{bmatrix} = \begin{bmatrix} a_{10} \\ a_{20} \\ a_{30} \\ a_{40} \end{bmatrix} + \sum_{l=1}^P \begin{bmatrix} a_{11}^l & a_{12}^l & a_{13}^l & a_{14}^l \\ a_{21}^l & a_{22}^l & a_{23}^l & a_{24}^l \\ a_{31}^l & a_{32}^l & a_{33}^l & a_{34}^l \\ a_{41}^l & a_{42}^l & a_{43}^l & a_{44}^l \end{bmatrix} \begin{bmatrix} MKT_{t-1} \\ WOM_{t-1} \\ SOCIAL_{t-1} \\ VALUE_{t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \\ e_{4t} \end{bmatrix}, \quad (3.1)$$

where  $MKT_t$  is the number of customer acquired through marketing actions at time  $t$ ;  $WOM_t$  is the number of customers acquired through word of mouth (referral) at time  $t$ ;  $SOCIAL_t$  is the number a customers acquired via social media advertising at time  $t$ ; and  $VALUE_t$  is the firm's financial performance at time  $t$ .  $P$  is model's lag order.

Using this approach the impact of various acquisition channels on company's performance can be directly compared. If it can be established that customers acquired through WOM and social media have significantly stronger impact on a company's financial performance than customers acquired by normal marketing activities, then

models can be built to investigate the acquisition process through social referral or social media.

#### *Acquisition through Social Referral*

The key to a successful referral system is: (1) identify customers with strong WOM power, who can help acquire new customers with high CLV; (2) decide incentive (reward) level offered to these customers for their referrals. The former is a profiling issue (of current customers), while the latter is an evaluation issue.

If a company has its customers' referral history, it can use a customers' past referral records to calculate their customer referral value (CRV) and identify customers with strong WOM power. For instance, in Kumar et al. (2007) the authors compare customers with high CLV and CRV. They find that customers with high CRV are not necessarily the customers with high CLV.

It would be interesting to further segment customers based on their social network characteristics and compare the referral results. It can be checked whether there is significant difference between more "sociable" customers and others.

In Kumar et al. (2010), the authors move forward to investigate the drivers behind CRV, thus make it possible to dynamically target new customers. In modeling CRV, Kumar et al. (2010) discuss the effect of various drivers, such as firm-specific factors, exchange characteristics (e.g., monetary value, frequency, cross-buy, product return policy), and customer characteristics (age and income). It will be interesting to include some social network characteristics into a CRV model and measure their effects. Such social network covariates can be characteristics such as network structure, social interactions, relationship strength, etc. These covariates are well

investigated in social network literature (e.g., Katona et al. 2011, Ansari et al. 2011). They can be useful to improve the prediction of CRV.

In Kumar et al. (2013), the authors use some interesting data collected from a social media campaign to measure customer influence effect (CIE) and customer influence value (CIV), both in monetary values. In particular, the authors estimate the drivers of CIE and CIV, including network and WOM characteristics (e.g., reciprocity, degree of transitive triangulation). The rationale for choosing these covariates and the method to estimate them can also be used to capture the drivers of customers' social referral power.

#### *Acquisition through Social Media*

More and more companies have added social media to their marketing mix (Armellini and Villanueva 2011, Weinberg and Pehlivan 2011). Compared with traditional marketing activities, social media marketing enables companies to have more direct, targeted, and personalized interactions with incumbent and prospect customers. Social media carries more credibility. It also enables companies to build a feedback loop between companies and customers. As a result, more and more companies are using social media to acquire new customers.

To acquire new customers with high CLV, a company needs to profile its current high value customers to identify their demographic characteristics (e.g., age, location) and behavioral patterns on social media (e.g., topic interested in, groups attended, "likes", etc.). Social media sites such as Facebook usually ask advertisers for this information to target specific customer segments and send messages to them ([www.facebook.com/ads](http://www.facebook.com/ads)).

Due to privacy protection concerns, collecting social information can be difficult. One possible solution is: (1) text mining high value customers' communications on company's platform, such as forums, discussion groups, blogs, etc. (2) crawling external social media sites to collect the information of high value customers' social behaviors. The former requires text mining capabilities and results can be quite accurate (e.g., Joachims 1999, Tausczik and Pennebaker 2010). The latter demands techniques to crawl on social website (Gjoka 2011). Only publicly available information can be collected.

After social behavior data is collected, researchers can use classification techniques to profile high value customers. On social media sites this profile information can be used to target the segment of promising potential customers.

### **3.4 Leverage the Power of Social Influence to Retain Customers**

Identifying and retaining high value customers is the key to a company's long term profitability. Various CRM papers have devoted to the topics such as customer status (e.g., Schmittlein et al. 1994, Allenby et al. 1999), CLV calculation (e.g., Gupta et al. 2006), and resource allocation (e.g., Reinartz et al. 2005). This section focuses on the impact of social influence on customer retention/defection within a social network.

#### *Social Influence on Customer Defection*

If the impact of one customer's churn on other customers' defection decisions can be directly modeled, influential customers in a social network can also be identified. If these influential customers choose to stay, customers connected with

them and under their influence are more likely to stay. If these influential customers choose to leave, then customers under their influence are more likely to leave as well. More resources should be allocated to retain these influential customers, so more customers can be retained through their social influence.

Research on this topic, however, is surprisingly sparse. One possible reason lies in the scantiness of data. In general, it is difficult to collect social network data and defection data at the same time. Not to mention that detecting customer churn is even harder, especially in a non-contractual setting.

Nitzan and Libai (2011) use hazard model and cell phone communication data to investigate whether there are social effects on customers' churn behaviors. They find that a customer is more likely to churn if customers communicating with her have churned.

In Chapter 2 this issue is addressed with a tie strength model and a defection decision model, jointly estimated with online gaming data. It is found that customers with frequent and recent interactions are more likely to have strong relationship. And customers with strong relationship have stronger impact on each other's defection decisions.

A lot of work remains to be done in this area, especially in the non-contractual setting. Nitzan and Libai (2011) discuss a contractual situation (cell phone service). In contrast, Chapter 2 models customer defection in a non-contractual scenario. Due to a unique situation in online video games, it is possible to detect a customer's true defection, even in a non-contractual setting. However, in reality most of the defection cases occur in non-contractual situations. Modeling defection in such situations

becomes very difficult because defection itself cannot be directly observed.

One solution is to use the approach proposed by Reinartz and Kumar (2000). In their paper the authors first use probability models to estimate the probability that a customer is still “alive”. They then calculate various expected values with estimated probabilities, and use these expected values as covariates to model customer’s life time value. This approach is also suitable for using hazard models to analyze customers’ length of relationship with companies.

Another solution is simply using some rule of thumb to decide whether a customer is still alive (e.g. a customer remaining inactive for a certain period is considered churned). This approach is widely used in industry (Reinartz and Venkatesan 2008). With this approach, the idle period itself can be considered as a covariate in the models as well. By jointly estimating this covariate an optimal judgment rule can be searched, along with the investigation of social impact on customers’ defection decisions.

#### *Satisfaction and Retention within a Social Network*

It has been well established that satisfied customers are more loyal than unsatisfied customers (e.g., Rust and Zahorik 1993, Bolton 1998, Gustafsson et al. 2005). However, it is still not very clear how customers influence each other’s satisfaction through their social connections. This is due to the difficulty of continuously monitoring customers’ satisfaction, let alone monitoring customer satisfaction within a social network.

Despite above difficulties, researchers have made several attempts. For example, Anderson (1998) points out that unsatisfied customers are more likely to



engage in WOM than satisfied customers. Fowler and Christakis (2008) find that happy people tend to cluster together within a social network.

If data is available, it will be interesting to investigate the mechanism through which customers influence each other's satisfaction. According to service quality theory (Oliver 1980, Rust and Oliver 1994), a customer's satisfaction/dissatisfaction comes from the gap between her expectation for the service and her actual perception of the service (e.g., Woodruff et al. 1983, Zeithaml et al. 1990, Boulding et al. 1993). Therefore, for each customer within a social network her expectation and perception can be modeled as a function of her own characteristics and her social ties' expectation and perception. Therefore, it is possible to capture the impact of WOM on customers' satisfaction.

#### *Service Failure and Recovery within a Social Network*

Due to intensifying competition, service failures could be devastating to service providers. Most recent examples include the service interruption of BlackBerry in 2011 and Netflix in 2012. With the growth of social network websites customers are becoming more closely connected. This makes the service failure issue even more critical. Negative word-of-mouth could spread very fast through social connections and a large scale customer defection could occur in the aftermath of a service failure. On the other hand, however, with appropriate service recovery actions, companies might be able to leverage the influence from social networks to retain their customers.

Because it is difficult to collect service failure data, so far most of the service failure/recovery literature uses survey, experiment, or simulation to investigate customers' reaction to service failures (e.g., Bolton 1998, McCollough et al. 2000, and

Zhu et al. 2004). Findings from these papers are similar: Excellent service recovery actions can raise customer satisfaction and increase re-patronage intentions. However, it is risky to completely rely on superior service recovery. The best strategy is to prevent service failures from happening.

So far no literature has combined service recovery research and social network analysis together. To do so, attempts can be made at group level and individual level.

First, customer groups' reaction to service failures can be modeled. These models can tell us whether groups with different network characteristics respond differently to service failures. For instance, it would be interesting to know whether a closely-knitted group is more likely to stay or churn. Regression models with covariates of network characteristics can be used.

Second, individual customer's reaction to service failures can be investigated with hazard models, such as the one proposed by Iyengar et al. (2011). One important purpose of individual level modeling is to identify influential customers. If influential customers can be identified, when a service failure occurs a company can allocate resources on these influential customers and persuade them to stay. As a result, customers connected with them will be more likely to stay.

#### *Social Network Value and CLV*

The heart of CRM is to identify and retain high value customers and generate more profits. High value customers can be identified with their CLV. Numerous papers have been published on estimating CLV and using CLV to optimize a company's CRM strategy. Most of these papers treat customers independently, ignoring the influence among connected customers. However, a customer with low

CLV might have strong influence on customers with high CLV. Apparently these customers are also valuable and deserve attention from the company. Hence, the CLV of a customer within a social network should have two components: an intrinsic value (the conventional CLV) and a social network value.

A customer's social network value can be defined as the expected lift on other customers' monetary contributions due this customer's social influence (Domingos and Richardson 2001). The estimation of social network value is not easy, because it is difficult to separate a customer's monetary contribution due to other customers' influence from the one due to her intrinsic reasons.

So far researchers have tried to investigate social network value in several specific scenarios. Gupta et al. (2009) measure the value of the direct network effects (buyer to buyer) and indirect network effects (buyer to seller) in a context of job posting website. Job seekers are welcome to register with the site for free ("free" customers). And because of the existence these job seekers recruiters are more willing to pay for the advertisements on the site. Therefore, "free" customers are also valuable.

Domingos and Richardson (2001) model customer's social network value in a setting of new product adoption. If a customer adopts a new product because another customer who is connect with her has adopted it, then the monetary contribution from this adoption can be considered as that influential customer's social network value.

In Chapter 2, an effort is made to measure a customer's social network value from her influence on other customers' defection decisions. First, customers' defection decisions within a social network are modeled. With this defection decision model the incremental impact of a customer's decision (stay/leave) on other customers' defection

decisions can be measured. If those customers churn, then all the expected monetary contribution from them will disappear. By combining the information of an influential customer's incremental impacts other customers' churn and the expected contributions from those customers; it is possible to measure the value of an influential customer's social influence. .

It still remains a challenge to measure customers' social network value in a more commonplace scenario such as repeated purchase. There is more work to be done in this area.

### **3.5 Grow Customers within a Social Network**

These days, growing customers has received more attention. Customer growth is defined as the expansion and growth of firm-customer relationships to increase both revenue and profits (Reinartz and Venkatesan 2008). With more customer data collected at individual level, it is possible to target individual customers, cross- and up-selling products to them (Reinartz and Venkatesan 2008).

This section discusses research opportunities in customer growth within a social network, focusing on three cross-selling scenarios: introducing brand-new products, promoting sequentially-purchased products, and increasing repeated purchase.

#### *Brand-New Products*

Introducing a brand-new product to a social network is indeed a new product diffusion/adoption issue. This issue has been well studied in social network literature. Two major categories of models have been used: diffusion models (e.g., Toubia et al.

2011) and linear models (e.g., Nair et al. 2010, Iyengar et al. 2011).

On the basis of these researches, the promotion strategies within a social network can be further analyzed. It would be interesting to know: (1) how influential customers react to the promotion messages; (2) how other customers adopt the new product under the influence from influential customers; (3) how to broadcast new product information to a social network. Are there any “hubs” in the network, which will facilitate the spread of the information about new products?

#### *Sequentially-Purchased Products*

Recently, most of the literature on the cross-selling of sequentially-purchased products has concentrated on financial services (e.g., Kamakura et al. 1991, Reinartz and Kumar 2000, Li et al. 2005). The main focus of these researches is to identify potential targets for direct marketing. Most researchers use customers’ own purchasing history to find customers who are more likely to buy the promoted products (e.g., Rossi et al. 1996, Knott et al. 2002, and Kumar et al. 2008).

If customers’ purchasing history and their social connections within a social network can be observed, a customer’s social connections’ purchasing history can also be used. Especially in the case of sequentially-purchased products, a customer’s friends’ purchasing pattern can serve as a good indicator for that customer’s purchasing intention.

This can be verified with the approach proposed by Kumar et al. (2008). Friends’ purchase patterns (e.g., the timing and quantity of purchases) might be a driver in the cross-buying model and the impact can be measured.

It could be interesting to investigate how to use friends’ purchase as a

promotion message. These days when a customer makes a purchase on a website, a button labeled “Like” or “Share with Friends” might appear on the webpage. If customer pushes the button, her friends will receive a message saying: “Your friend has purchased (or likes) this product”. Customers’ reaction to this type messages can be modeled and the effectiveness of messages on a customer’s purchase intention can be measured.

To begin with, lab experiments or field experiments can help observe customers’ reaction to these messages. If these messages are effective then modeling approach can be used to quantitatively gauge the effectiveness. As mentioned by Venkatesan et al. (2007), investigation can start with checking the existence of a “U-shape” relationship between the amount of information sent and the purchase intention of customers who receive the information. Firms can use the findings to optimize their promotion strategy.

### *Repeat Purchase*

Modeling social influence in a repeat purchase setting is a challenge, because it is difficult to “filter out” the impact of social influence. Also, unlike the “one-to-one” influence discussed in Trusov et al. (2010) and Iyengar et al. (2011) (e.g., an opinion leader’s influence on her followers), in a setting of repeat purchase “group-to-one” influence is more common. In other words, when a customer is considering a purchase, she is immersed in the influence from a group of customers who are connected to her.

As an attempt to address this issue, in Chapter 1 the impact of group-to-one social influence on each customer’s repeat purchases is modeled with the data collected from an online video game. In particular, the influence from strong ties and

weak ties in the network is compared. It is found that as a whole, weak ties are more influential than strong ties. This is because: (1) weak ties are much more widespread than strong ties; (2) information (especially new information) disseminates much faster through weak ties than through strong ties.

The impact of product sharing on customers' purchasing behaviors is also investigated. It is found that if sharing is allowed, customers are more willing to buy new products. This is because customers are more willing to take risks when they know that if they are not satisfied with the product they can exchange it with their friends.

The biggest obstacle in this area is still data collection. In general, it is very difficult to collect data on both social connections/interactions and purchases. So far the only data sources used by researchers have been collected from online gaming and telecommunication. With more data available, more researches will be done.

### **3.6 Capture Firm-Customer Relationship Dynamics within a Social Network**

The ongoing relationship between a firm and its customers evolves over time. The relationship state depends on customers' own characteristics, firm's CRM activities, and other customers' influence if they are connected within a social network. To investigate the dynamics of firm-customer relationship state, hidden Markov models (HMMs) are widely used.

With HMM, firm-customer relationship is modeled as a latent state (Pfeifer 2000). At time  $t$ , each customer stays in state  $s$  ( $s = 1, \dots, NS$ ) with a probability. Customer  $i$  starts with initial probabilities described in vector  $\boldsymbol{\pi}_i = [\pi_{i1}, \dots, \pi_{iNS}]'$ .

Starting with initial distribution, customer  $i$ 's state evolves as a Markov chain . The probabilities of transition across states are described in transition matrix  $\mathbf{Q}_{i,t-1 \rightarrow t} = \{q_{i,s_{t-1},s_t}\}$ , where  $q_{i,s_{t-1},s_t}$  is the probability that customer  $i$  shifts from state  $s_{t-1}$  at time  $t - 1$  to state  $s_t$  at time  $t$ . Transition probabilities can be homogeneous (not time-varying) or nonhomogeneous (time-varying). In most CRM cases, transition probabilities are modeled as a function of time-varying covariates. By doing so, researchers can investigate the effectiveness of a firm's CRM efforts intended to enhance its relationship with customers.

Because relationship states are latent, researchers can only observe customers' behaviors, which depend on the relationship states. The distribution of customer  $i$ 's state-dependent behaviors is described by a state-dependent distribution  $\mathbf{m}_{it} = [m_{it|1}, \dots, m_{it|NS}]'$ , where  $m_{it|s}$  is the probability of customer  $i$  taking certain action when she is in state  $s$ . Thus the three key components of a HMM are: (1) initial state distribution, (2) transition matrix, and (3) state-dependent distribution. If customer  $i$ 's behaviors from time 1 to time  $T$ ,  $y_{i1}, \dots, y_{iT}$ , can be observed, then the likelihood can be calculated with the following equation (for details, see Netzer et al. 2008):

$$L_{iT} = P_i(Y_{i1} = y_{i1}, \dots, Y_{iT} = y_{iT}) = \sum_{s_1=1}^{NS} \sum_{s_2=1}^{NS} \dots \sum_{s_T=1}^{NS} \left[ \pi_{is_1} \prod_{\tau=2}^T q_{is_\tau s_{\tau-1}} \cdot \prod_{\tau=1}^T m_{(i\tau|s_\tau)}^{y_{i\tau}} (1 - m_{(i\tau|s_\tau)})^{1-y_{i\tau}} \right]. \quad (3.2)$$

Apparently, it is difficult to calculate this likelihood. Therefore various algorithms have been developed to estimate HMMs. The most commonly used



algorithms include Expectation-Maximization (EM) and Monte Carlo Markov Chain (MCMC). MCMC algorithm is particularly important when the state transition is modeled as nonhomogeneous. Also, if we want to capture heterogeneity among customers, a hierarchical Bayesian model is usually used (e.g., Netzer et al. 2008), which is usually estimated with MCMC. Fortunately, if customers can be assumed independent with each other, MacDonald and Zucchini (1997, 2009) propose a highly simplified and tractable likelihood specification:

$$L_{iT} = P_i(Y_{i1} = y_{i1}, \dots, Y_{iT} = y_{iT}) = \boldsymbol{\pi}_i \tilde{\mathbf{m}}_{i1} \mathbf{Q}_{i,1 \rightarrow 2} \cdots \mathbf{Q}_{i,T-1 \rightarrow T} \tilde{\mathbf{m}}_{iT} \mathbf{1}', \quad (3.3)$$

where  $\boldsymbol{\pi}_i$  is the initial distribution of customer  $i$ 's state;  $\mathbf{Q}_{i,1 \rightarrow 2}, \dots, \mathbf{Q}_{i,T-1 \rightarrow T}$  are customer  $i$ 's transition matrices,  $\tilde{\mathbf{m}}_{it}$  is an  $NS \times NS$  diagonal matrix with state-dependent distribution vector  $\mathbf{m}_{it}$  on its diagonal. In their paper, Netzer et al. (2008) demonstrate how to use HMM to model the dynamic relationship between a school and its alumni. As an extension of this research, Montoya et al. (2010) use HMM to capture the relationship dynamics in a setting of new drug introduction. They then use dynamic programming to optimize the allocation of detailing resources.

So far, in all HMM literature researchers treat customers independently, and the likelihood specification given in Zucchini and MacDonald (1997, 2009) is particularly popular among researchers. Tractable and simple, with this specification likelihood can be calculated with standard matrix computation.

Unfortunately, this assumption of independence might no longer be applicable within a social network. For instance, it is quite possible that a customer's relationship

with the firm is the function of her friends' relationships with the firm. One customer's transition probability among relationship states might depend on her friends' relationship states as well.

However, it is extremely difficult to specify the likelihood function when customers are no longer independent from each other. Researchers in computer science and management have been working hard to solve this problem. But so far no progress has been reported. Despite the technical difficulties, this topic is still a very interesting and carries significant managerial implications.

### **3.7 Conclusion**

Researches in customer relationship management and social network analysis have been undergoing fast growth. With increasing data availability, time is ripe for researchers to investigate the intersection of these two disciplines. The combination of CRM and SNA bring forth tremendous research opportunities. Three research areas merit researchers' consideration: (1) quantitatively measuring the social influence in each aspect of CRM; (2) managing customers as a network; (3) capturing the dynamics of firm-customer and customer-customer relationships.

#### *Measuring Social Influence*

Scenarios and methodology to measure social influence vary from one aspect of CRM to another. For instance, in customer adoption aspect, the focus is to measure the referral power of incumbent customers from their social behaviors, thus identify influential customers and encourage them to help acquire new customers through social media and other channels.

In customer retention aspect, it is interesting to understand how customers influence each other's expectation and perception of service and products. It is also critical to model customer churn within a network, thus companies can leverage the social influence to retain their customers.

In customer growth aspect, it is important to measure the social influence among customers in terms of their purchasing behaviors. Particularly, it will be interesting to investigate the mechanism of social influence on customers' repeat purchase. It will also be interesting to explore how to use customers' social interaction and purchasing history to identify cross-selling opportunities.

#### *Managing Customers as Networks*

Based on a thorough understanding of how customers influence each other within a network, companies can treat customer groups as the units for their CRM efforts. This will first benefit a company's communication strategy. For instance, if a company wants to promote some products to its customers, it can send the message to the "hubs" in a customer network and encourage them to disseminate the information. The remaining work will be done by the WOM.

It will also be interesting to track the growth of customer networks. CLV of various customer groups can be measured and tracked. The distribution of CLV within the network can also be analyzed. Knowing the impact of network characteristics on customer groups' value and growth would help a company find its optimal "social network strategy".

#### *Capture Relationship Dynamics in a Social Network*

It is challenging to model the dynamic relationships between a company and

its customers within a social network. However, if how customers' relationship states are dependent on each other can be modeled, researchers will have a deeper understanding of the influence mechanism within a social network. This will require a breakthrough in methodology, but it is definitely a rewarding direction.

In sum, a thorough understanding of social influence mechanism among customers will enable companies to leverage the power of social network to acquire, retain and grow high value customers, and enhance firm-customer and customer-customer relationships. This cross-disciplinary area deserves more attention from marketing researchers.

## REFERENCES

- Alba, Richard D. (1982), "Taking Stock of Network Analysis: A Decades Results," *Research in the Sociology of Organizations*, 1: 39-74.
- Allenby, Greg M., Robert P. Leone and Lichung Jen (1999), "A Dynamic Model of Purchase Timing with Application to Direct Marketing," *Journal of the American Statistical Association*, 94(446), 365-374.
- Ansari, Asim, Oded Koenigsberg and Florian Stahl (2011), "Modeling Multiple Relationships in Social Networks," *Journal of Marketing Research*, 48(August) 713-728.
- Aral, Sinan and Dylan Walker (2011), "Creating Social Contagion Through Viral Product Design," *Management Science*, 57(9), 1623–1639.
- Armellini, Guillermo and Julian Villanueva (2011), "Adding Social Media to the Marketing Mix," *IESE-Insight Magazine*, 3(4), 29-36.
- Bass, Frank M. (1969), "A New Product Growth for Model Consumer Durables," *Management Science*, 15(5), 215–227.
- Bass, Frank M., Trichy V. Krishnan and Dipak C. Jain (1994), "Why the Bass Model Fits without Decision Variables," *Marketing Science*, 13(3), 203-223.
- Berger, Jonah and Eric M. Schwartz (2011), "What Drives Immediate and Ongoing Word of Mouth?" *Journal of Marketing Research*, 48(October), 869-880.
- Biyalogorsky, Eyal, Eitan Gerstner and Barak Libai (2001), "Customer Referral Management: Optimal Reward Programs," *Marketing Science*, 20(1), 82-95.
- Bodapati, Anand and Sachin Gupta (2004), "A Direct Approach to Predicting discretized Response in Target Marketing," *Journal of Marketing Research*, 41(February), 73-85.
- Bolton, Ruth N. (1998), "A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: The Role of Satisfaction," *Marketing Science*, 17(1) 45–65.
- Bult, Jan R. and Tom Wansbeek (1995), "'Optimal Selection for Direct Mail,'" *Marketing Science*, 14(4), 378-394.

Carrington, Peter J., John Scott and Stanley Wasserman, eds., *Models and Methods in Social Network Analysis*, Cambridge University Press, 2005.

Domingos, Pedro, and Matt Richardson (2001), "Mining the Network Value of Customers," In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2001.

Dover, Yaniv, Jacob Goldenberg and Daniel Shapira (2012), "Uncovering Degree Distribution from Adoption Data," *Marketing Science*, (2012), 31(4), 689–712.

Du, Rex Yuxing and Wagner A. Kamakura (2011), "Measuring Contagion in the Diffusion of Consumer Packaged Goods," *Journal of Marketing Research*, 48(February), 28-47.

Fowler, James H. and Nicholas A. Christakis (2008), "Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis Over 20 Years in the Framingham Heart Study," *BMJ: British medical journal* 337.

Gjoka, Minas, Maciej Kurant, Carter T. Butts and Athina Markopoulou (2011), "Practical Recommendations on Crawling Online Social Networks," *IEEE Journal on Selected Areas in Communications*, 29(9), 1872-1892.

Godes, David and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23(4), 545-560.

Gupta, Sunil, Dominique Hanssens, Bruce Hardie, William Kahn, V. Kumar, Nathaniel Lin, Nalini Ravishanker and S. Sriram (2006), "Modeling Customer Lifetime Value," *Journal of Service Research*, 9(2), 139-155.

Gupta, Sunil and Donald R. Lehmann (2008), "Models of Customer Value," *Handbook of Marketing Decision Models*, Springer Science + Business Media, LLC.

Gupta, Sunil, Donald R. Lehmann and Jennifer A. Stuart (2004), "Valuing Customers," *Journal of Marketing Research*, 41(February), 7-18.

Gupta, Sunil, Carl F. Mela, and Jose M. Vidal-Sanz (2009), "The Value of a "Free" Customer," *Working Paper*.

Gupta, Sunil and Valarie Zeithaml (2006), "Customer Metrics and Their Impact on Financial Performance," *Marketing Science*, 25(6), 718-739.

Gustafsson, Anders, Michael D. Johnson and Inger Roos (2005), "The Effects of Customer Satisfaction, Relationship Commitment Dimensions, and Triggers on Customer Retention," *Journal of Marketing*, 69(4) 210–218.

Hebden, J. J., and J. F. Pickering (1974), "Patterns of Acquisition of Consumer Durables," *Oxford Bulletin of Economics and Statistics*, 36(2), 67-94.

Hogan, John E., Donald R. Lehmann, Maria Merino, Rajendra K. Srivastava, Jacquelyn S. Thomas and Peter C. Verhoef (2002), "Linking Customer Assets to Financial Performance," *Journal of Service Research*, 5(1), 26-38.

Iacobucci, Dawn (1996), *Networks in Marketing*, Thousand Oaks: Sage Publications.

Iacobucci, Dawn and Nigel Hopkins (1992), "Modeling Dyadic Interactions and Networks in Marketing," *Journal of Marketing Research*, 29(February), 5-17.

Iyengar, Raghuram, Christophe Van den Bulte and Thomas W. Valente (2011), "Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science*, 30(2), 195-212.

Joachims, Thorsten (1999), "Transductive Inference for Text Classification Using Support Vector Machines," in *ICML*, vol. 99, pp. 200-209.

Kamakura, Wagner A., Michel Wedel, Fernando de Rosa and Jose Afonso Mazzon (2003), "Cross-Selling through Database Marketing: A Mixed Data Factor Analyzer for Data Augmentation and Prediction," *International Journal of Research in Marketing*, 20(1), 45-65.

Kamakura, Wagner A., Sridhar N. Ramaswami and Rajendra K. Srivastava (1991), "Applying Latent Trait Analysis in the Evaluation of Prospects for Cross-Selling of Financial Services," *International Journal of Research in Marketing*, 8(4), 329-349.  
Katona, Zsolt, Peter P. Zubcsek and Miklos Sarvary (2011), "Network Effects and Personal Influences: The Diffusion of an Online Social Network," *Journal of Marketing Research*, 48(June), 425-443.

Kim, Namwoon, Vijay Mahajan and Rajendra K. Srivastava (1995), "Determining the Going Market Value of a Business in an Emerging Information Technology Industry: The Case of the Cellular Communications Industry," *Technological Forecasting and Social Change*, 49(3), 257-279.

Knott, Aaron, Andrew Hayes, Scott A. Neslin (2002), "Next-product-to-buy Models for Cross-Selling Applications," *Journal of Interactive Marketing*, 16(3), 59-75.

Kumar, V., Vikram Bhaskaran, Rohan Mirchandani and Milap Shah (2013), "Practice Prize Winner-Creating a Measurable Social Media Marketing Strategy: Increasing the Value and ROI of Intangibles and Tangibles for Hokey Pokey," *Marketing Science*, 32(2), 194-212.

Kumar, V., J. Andrew Petersen, and Robert P. Leone (2007), "How Valuable Is Word of Mouth?" *Harvard Business Review*, 85(10), 139-46.

- Kumar, V., J. Andrew Petersen, Robert P. Leone (2010), "Driving Profitability by Encouraging Customer Referrals: Who, When, and How," *Journal of Marketing*, 74(5), 1-17.
- Kumar, V., Rajkumar Venkatesan and Werner Reinartz (2008), "Performance Implications of Adopting a Customer-Focused Sales Campaign," *Journal of Marketing*, 72(5), 50-68.
- Lee, Jae Young and David R. Bell (2013), "Neighborhood Social Capital and Social Learning for Experience Attributes of Products," *Marketing Science*, forthcoming.
- Lemon, Katherine N., Tiffany Barnett White and Russell S. Winer (2002), "Dynamic Customer Relationship Management: Incorporating Future Considerations into the Service Retention Decision," *The Journal of Marketing*, 66(January), 1-14.
- Li, Shibo, Baohong Sun and Ronald T. Wilcox (2005), "Cross-Selling Sequentially Ordered Products: An Application to Consumer Banking Services," *Journal of Marketing Research*, 42(May), 233-239.
- Libai, Barak, Eitan Muller and Renana Peres (2009), "The Diffusion of Services," *Journal of Marketing Research*, 46(April), 163-175.
- Libai, Barak, Eitan Muller and Renana Peres (2013), "Decomposing the Value of Word-of-Mouth Seeding Programs: Acceleration versus Expansion," *Journal of Marketing Research*, 50(April), 161-176.
- Lovejoy, William S. and Amitabh Sinha (2010), "Efficient Structures for Innovative Social Networks," *Management Science*, 56(7), 1127-1145.
- Malthouse, Edward C. (2001), "Assessing the Performance of Direct Marketing Scoring Models," *Journal of Interactive Marketing*, 15(1), 49-62.
- Mallapragada, Girish, Rajdeep Grewal and Gary Lilien (2012), "User-Generated Open Source Products: Founder's Social Capital and Time to Product Release," *Marketing Science*, 31(3), 474-492.
- Manchanda, Puneet, Ying Xie and Nara Youn (2008), "The Role of Targeted Communication and Contagion in Product Adoption," *Marketing Science*, 27(6), 961-976
- McCollough, Michael A., Leonard L. Berry, and Manjit S. Yadav (2000), "An Empirical Investigation of Customer Satisfaction After Service Failure and Recovery", *Journal of Service Research*, 3(2), 121-137.



Mertz, Sharon A (2009), "Dataquest Insight: CRM Software Market Share Analysis, Worldwide, 2008," Gartner Inc., July 6.

Montoya, Ricardo, Oded Netzer and Kamel Jedidi (2010), "Dynamic Allocation of Pharmaceutical Detailing and Sampling for Long-term Profitability," *Marketing Science*, 29(5), 909-924.

Moreno, Jacob L. (1934), "Who Shall Survive?: A New Approach to the Problem of Human Interrelations," *Nervous and Mental Disease Monograph Series*, no 58.

Nair, Harikesh S., Puneet Manchanda and Tulikaa Bhatia (2010), "Asymmetric Social Interactions in Physician Prescription Behavior: The Role of Opinion Leaders," *Journal of Marketing Research*, 47(October), 883-895.

Nam, Sungjoon, Puneet Manchanda and Pradeep K. Chintagunta (2010), "The Effect of Signal Quality and Contiguous Word of Mouth on Customer Acquisition for a Video-on-Demand Service," *Marketing Science*, 29(4), 690-700.

Narayan, Vishal, Vithala R. Rao and Carlyne Saunders (2011), "How Peer Influence Affects Attribute Preferences: A Bayesian Updating Mechanism," *Marketing Science*, 30(2), 368-384.

Neslin, Scott A., Sunil Gupta, Wagner Kamakura, Junxiang Lu, and Charlotte H. Mason (2006), "Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models," *Journal of Marketing Research*, 43 (May), 204-211.

Netzer, Oded, James M. Lattin and Vikram Srinivasan (2008), "A Hidden Markov Model of Customer Relationship Dynamics," *Marketing Science*, 27(2), 185-204.

Oliver, Richard L. (1980), "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions," *Journal of Marketing Research*, 17(November), 46-69.

Payne, Adrian and Pennie Frow (2005), "A Strategic Framework for Customer Relationship Management," *Journal of Marketing*, 60 (October), 167-176.

Parasuraman, A. (2006), "Invited Commentary Modeling Opportunities in Service Recovery and Customer-Managed Interactions," *Marketing Science*, 25(6), 590-593.

Pfeifer, Phillip E. and Robert L. Carraway (2000), "Modeling Customer Relationships as Markov Chains," *Journal of Interactive Marketing*, 14(2), 43-55.

Ransbotham, Sam, Gerald C. Kane and Nicholas H. Lurie (2012), "Network Characteristics and the Value of Collaborative User-Generated Content," *Marketing Science*, 31(3), 387-405.

Reinartz, Werner J. and Vijay Kumar (2000), "On the Profitability of Long-life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing," *The Journal of Marketing*, 64(4), 17-35.

Reinartz, Werner J. and Vita Kumar (2003), "The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration," *Journal of marketing*, 67(1), 77-99.

Reinartz, Werner, Jacquelyn S. Thomas and V. Kumar (2005), "Balancing Acquisition and Retention Resources to Maximize Customer Profitability," *Journal of Marketing*, 69(1), 63-79.

Reinartz, Werner J. and Rajkumar Venkatesan (2008), "Decision Models for Customer Relationship Management (CRM)," *Handbook of Marketing Decision Models*, Springer Science + Business Media, LLC.

Rossi, Peter E., Robert E. McCulloch and Greg M. Allenby (1996), "The Value of Purchase History Data in Target Marketing," *Marketing Science*, 15(4), 321-340.

Rust, Roland T. and Tuck Siong Chung (2006), "Marketing Models of Service and Relationships," *Marketing Science*, 25(6), 560-580.

Rust, Roland T. and Richard L. Oliver (1994). *Service Quality: New Directions in Theory and Practice*. London: Sage.

Rust, Roland T. and Anthony J. Zahorik (1993), "Customer Satisfaction, Customer Retention, and Market Share," *Journal of retailing*, 69(2): 193-215.

Schmitt, Philipp, Bernd Skiera, Christophe Van den Bulte (2011), "Referral Programs and Customer Value," *Journal of Marketing*, 75(1), 46-59.

Schmittlein, David C. and Robert A. Peterson (1994), "Customer Base Analysis: An Industrial Purchase Process Application," *Marketing Science*, 13(1), 41-67.

Scott, John and Peter J. Carrington, eds., *The SAGE Handbook of Social Network Analysis*, SAGE Publications, 2011.

Schmittlein, David C. and Robert A. Peterson (1994), "Customer Base Analysis: An Industrial Purchase Process Application," *Marketing Science*, 13(1), 41-67.

Shriver, Scott K., Harikesh S. Nair and Reto Hofstetter (2013), "Social Ties and User-Generated Content," *Management Science*, 59(6), 1425-1443.

Smith, Amy K., Ruth N. Bolton and Janet Wagner (1999), "A Model of Customer Satisfaction with Service Encounters Involving Failure and Recovery," *Journal of Marketing Research*, 36 (August), 356-372.

Spekman, Robert E. (1996), "A Reflection on Two Decades of Business-to-Business Marketing Research - Implications for Understanding Marketing Relationships and Networks," in Dawn Iacobucci (eds.), *Networks in Marketing* (pp. 8-15), Thousand Oaks: Sage Publications.

Steenburgh, Thomas J., Andrew Ainslie and Peder Hans Engebretson (2003), "Massively Categorical Variables: Revealing the Information in Zip Codes," *Marketing Science*, 22(1), 40-57.

Stephen, Andrew and Olivier Toubia (2010), "Deriving Value from Social Commerce Networks," *Journal of Marketing Research*, 47(April), 215-228.

Tausczik, Yla R. and James W. Pennebaker (2010), "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods," *Journal of Language and Social Psychology*, 29(1), 24-54.

Thomas, Jacquelyn S., Robert C. Blattberg and Edward J. Fox (2004), "Recapturing Lost Customers," *Journal of Marketing Research*, 41(February), 31-45.

Toubia, Olivier, Jacob Goldenberg and Rosanna Garcia (2009), "Diffusion Forecasts Using Social Interactions Data," *MSI Special Report*, 09-210.

Trusov, Michael, Anand V. Bodapati and Randolph E. Bucklin (2010), "Determining Influential Users in Internet Social Networks," *Journal of Marketing Research*, 47(August), 643-658.

Tucker, Catherine (2008), "Identifying Formal and Informal Influence in Technology Adoption with Network Externalities," *Management Science*, 54(12), 2024-2038.

Venkatesan, Rajkumar, Vipin Kumar and Nalini Ravishanker (2007), "Multichannel Shopping: Causes and Consequences," *Journal of Marketing*, 71(April), 114-132.

Verhoef, Peter C. and Bas Donkers (2005), "The Effect of Acquisition Channels on Customer Loyalty and Cross-buying," *Journal of Interactive Marketing*, 19(2), 31-43.

Villanueva, Julian, Shijin Yoo and Dominique M. Hanssens (2008), "The Impact of Marketing-Induced vs. Word-of-Mouth Customer Acquisition on Customer Equity Growth," *Journal of Marketing Research*, 45(February), 48-59

Weinberg, Bruce D. and Ekin Pehlivan (2011), "Social Spending: Managing the Social Media Mix," *Business Horizons*, 54(3), 275-282.

Woodruff, Robert B., Ernest R. Cadotte and Roger L. Jenkins (1983), "Modeling Consumer Satisfaction Processes using Experience-Based Norms," *Journal of Marketing Research*, 20 (August), 296-304.

Yoganarasimhan, Hema (2012), “Impact of Social Network Structure on Content Propagation: A Study Using YouTube Data,” *Quantitative Marketing and Economics*, 10:111–150

Zeithaml, V.A., A. Parasuraman, and L.L. Berry (1990), *Delivering Quality Service*, The Free Press, New York

Zhu, Zhen, K. Sivakumar, and A. Parasuraman (2004), “A Mathematical Model of Service Failure and Recovery Strategies”, *Decision Sciences*, 35(3), 493-525.